

A entoação em falas teatrais: uma análise da raiva e da fala neutra

(The intonation in play-acted vocal emotions: analysis of anger and neutral speech)

Aline Mara De Oliveira Vassoler¹, Marcus Vinícius Moreira Martins²

¹Departamento de Estudos Linguísticos (Oralidade e Letramento) – Universidade Estadual Paulista (Unesp – Rio Preto)

²Departamento de Letras Clássicas e Vernáculas da Faculdade de Filosofia, Letras e Ciências Humanas e Instituto de Matemática e Estatística – Universidade de São Paulo (USP)

alineoliveiravassoler@gmail.com, marcusmartins@usp.br

Abstract: In order to investigate emotions in speech, specifically anger, three professional actresses (20 years in the profession) participated in the survey. The play-acted speeches were captured in the laboratory of the Studio Multimedia Electronic Computer Center of the University of São Paulo (CCE-USP), using Sound Forge 9 audio software. The software used to obtain the curve intonational average of three repetitions of anger and neutral speech was the routine ExProsodia. The average of F_0 mean obtained for rage was 197 Hz and for neutral speech was 156 Hz. Anger behaves in high frequency bands, while the F_0 of neutral speech is located in medium frequency bands. However, the correlation was 0.88 for both conditions, showing that the linear range is nearly identical.

Keywords: emotion; intonation; experimental phonetics; acoustical analysis.

Resumo: Para investigar as emoções na fala, especificamente a raiva, três atrizes profissionais (20 anos de profissão) participaram da pesquisa. A captura das vozes foi realizada no laboratório Estúdio Multimeios do Centro de Computação Eletrônica da Universidade de São Paulo (CCE-USP), captadas pelo *software* de edição de som *Sound Forge 9*. O programa utilizado para obtermos a curva entoacional média das três repetições da raiva e da fala neutra foi a rotina *ExProsodia*. A média das médias de F_0 obtida para a raiva foi de 197 Hz e para a fala neutra, 156 Hz. A raiva se comporta nas faixas de altas frequências, enquanto a F_0 da fala neutra está localizada nas faixas de médias frequências. No entanto, a correlação foi de 0,88 para ambas as situações, mostrando que a variação linear é quase idêntica.

Palavras-chave: emoção; entoação; fonética experimental; análise acústica.

Introdução

As emoções humanas são transmitidas por meio das expressões faciais, dos gestos corporais e pela fala do indivíduo. É evidente que a compreensão da fala depende diretamente de informações sintáticas e semânticas. No entanto, pistas prosódicas também carregam conteúdos fundamentais para a compreensão da mensagem. Essas informações permitem que o ouvinte, não só codifique e decodifique toda a mensagem falada, como também compreenda importantes funções linguísticas como a (i) personalidade do indivíduo, (ii) a idade do falante, (iii) o sexo/gênero e (iv) a condição física e emocional do falante (MOZZICONACCI, 2002; MOZZICONACCI; HERMES, 1999). Os elementos paralinguísticos e/ou não verbais, como as expressões faciais, os gestos, os sons (como os suspiros), também estão inerentes à compreensão da mensagem (MOZZICONACCI,

2001). Os primeiros estudos envolvendo as emoções (DARWIN, 1872), já consideravam as mudanças vocais como parâmetro de análise das emoções, no entanto, o foco principal desses estudos, eram as expressões faciais das diferentes emoções.

A partir do século XIX, as pesquisas sobre a fala se intensificaram e as relações entre as emoções e as mudanças vocais passaram a ser investigadas. Os estudos deste período, envolvendo o sinal acústico, comprovaram as primeiras inferências impressionísticas sobre as relações entre a voz e as emoções. Essas pesquisas demonstraram que as emoções geram mudanças no padrão respiratório, na fonação e na articulação do aparelho fonador do indivíduo, alterando, assim, os parâmetros do sinal acústico (SCHERER, 1989; WILLIAMS; STEVENS, 1972; BANSE; SCHERER, 1996). Dessa maneira, diversos experimentos têm confirmado a possibilidade de diferenciar as emoções por meio de parâmetros acústicos (MURRAY; ARNOTT, 1993, VASSOLER, 2011; OLIVEIRA; MEDEIROS, 2010).

O objetivo deste trabalho é contribuir com este corpo de pesquisa, analisando a emoção da raiva por meio de análises de falas interpretadas por atrizes e pela fala neutra (melhores definidas na seção a seguir). Pretendemos, pois, entender se há padrões entoacionais nas curvas entoadas com as emoções acima apresentada. Nossa proposta, é analisar as curvas entoacionais da fala emotiva da raiva comparando-a com a fala neutra, utilizando-se do modelo teórico metodológico desenvolvido por Ferreira Netto (2006, 2008), bem como realizar uma análise automática usando a rotina ExProsodia.

A Entoação

O fenômeno da entoação passou a ser melhor compreendido em suas funções linguísticas e paralinguísticas a partir de meados de década de 70. Neste período surge um grande número de publicações em torno do tema, tais como as teses de Libermann (1975), Maeda (1976) e Pierrehumbert (1980) e os trabalhos do IPO publicados por t'Hart (1967), t'Hart e Cohen (1973), t'Hart (1981) e reunidos em Collier, t'Hart e Cohen (1990).

Esta geração de pesquisadores preocupou-se, principalmente, em definir e caracterizar linguisticamente o fenômeno da entoação. Para este fim, assumiram um pressuposto teórico-metodológico inicial: a entoação deve ser compreendida, apenas como a variação de F_0 . Tal definição, embora operacional, incluía ainda os tons lexicais, como aqueles que ocorrem nas línguas da família *bantu* ou no chinês, por exemplo. Por esta razão, uma segunda definição teve que ser assumida: a entoação é pós-lexical, não sendo intrínseca aos itens do léxico. Ainda que se note, em alguns casos, tons cristalizados em itens lexicais, como no marcador conversacional “né”, ao qual se atribui uma curva ascendente. A terceira definição ainda nos diz que a entoação transmite informações complementares como: “tipo de sentença ou ato de fala, ou foco e estrutura da informação” (LADD, 1996, p. 7). Ladd (1996) ainda completa “por esta definição, entoação exclui características como acento tonal e tom, os quais são determinados pelo léxico e que servem para distinguir uma palavra da outra.”.

A terceira definição assume que pode haver dois tipos de informações veiculadas pela entoação, sendo uma linguística e outra paralinguística. A informação linguística, diz respeito a aspectos especificamente linguísticos, como a asserção ou a interrogação,

ao passo que a informação paralinguística transmitiria informações suplementares que contribuem para a compreensão da linguagem de modo geral (LADD, 1996). A hipótese central é a de que a primeira (informação linguística) pode ser estruturada em termos de uma teoria ou metodologia que se preocupe unicamente com questões associadas à entoação e prosódia, como proposto pelo programa ToBI (SILVERMAN et al., 1992; JUN, 2006). Por outro lado, a informação paralinguística não pode ser organizada ou explicada em termos de uma teoria linguística e demandaria outros modelos de análise.

Pike (1945) ao discutir a cita diferença, propôs que haveria dois canais de informação sobrepostos. Um primeiro propriamente linguístico e um segundo corpóreo, por assim dizer. O primeiro está diretamente associado ao código linguístico, isto é, aos segmentos, a fonologia, sintaxe, entre outros. Enquanto o segundo está ligado aos gestos corporais, a face do interlocutor e outros aspectos corporais. Desta maneira, é a associação entre os dois canais que fornece a informação paralinguística. Como apresentado em Pike (1945): um insulto falado por uma face risonha, pode ser compreendido como um elogio, por exemplo.

Ladd (1996) aponta que embora estes aspectos tenham influência no que é entendido pelo falante, eles não alteram a forma fonológica do que é dito. Nosso trabalho se sustenta nesta afirmação, pois é a partir dela que podemos comparar uma mesma sentença, a qual é entoada de diversas maneiras. A forma fonológica nos garante que teremos uma mesma representação mental, ou alvo tonal (XU; WANG, 2001), o qual será realizado pelo falante. Embora, nesta realização possamos ter valores de F_0 diferentes, seja intra-sujeitos, seja entre-sujeitos.

Prosódia e emoção

Vaissière (2004) e Hirst e DiCristo (1998) discutem, em seus respectivos trabalhos, o fato de não haver uma distinção clara entre entoação e prosódia. Por esta razão, os autores assumem duas definições de prosódia, uma *estreita* e outra *ampla*. Na abordagem estreita, os fenômenos de F_0 são tomados como base para a análise, seja ela linguística ou não, ao passo que na abordagem ampla outros fenômenos como a intensidade, a qualidade vocal, o ritmo, acento, entre outros, podem ser analisados. Os estudos linguísticos tem se pautado por uma análise restrita, isto é, os demais aspectos prosódicos não são analisados em conjunto, mas, sim, como parte do processo de produção (PIERREHUMBERT, 1980; T'HART; COLLIER; COHEN, 1990; FERREIRA NETTO, 2006, 2008), enquanto os estudos sobre as emoções tem se voltado para uma abordagem ampla, como veremos a seguir.

Os principais sinais acústicos analisados para a expressão vocal das emoções são: (i) o contorno da frequência fundamental; (ii) a variação de F_0 e (iii) intensidade vocal. As principais técnicas de análise utilizadas são: (i) o cálculo da distribuição de energia no espectro de frequência (especialmente a energia que envolve as regiões de alta e de baixa frequência e que afeta a percepção da qualidade de voz ou do timbre); (ii) a localização dos formantes ($F1$ e $F2$ relacionados com a percepção da articulação) e (iii) variedade temporal dos fenômenos, incluindo durações e pausas (BANSE; SCHERER, 1996; MURRAY; ARNOT, 1993).

No entanto, Scherer et al. (1984) e Murray e Arnott (1993) comprovaram que a frequência fundamental F_0 parece ser o parâmetro acústico mais importante na identificação

das emoções na fala. A entonação da fala é representada acusticamente pela variação de F_0 (LADEFOGED, 1974), por isso, pode ser utilizada na diferenciação das emoções (WILLIAMS; STEVENS, 1972). A maioria das pesquisas que estudam os efeitos vocais das emoções são adeptas da teoria discreta das emoções, que estabelece entre 9 e 14 emoções como as emoções básicas ou fundamentais. Essas emoções caracterizam-se por apresentarem padrões de respostas fisiológicas muito específicas, tanto na expressão facial, quanto na vocal. Assim, as emoções consideradas básicas são a alegria, a tristeza, a raiva, o medo e a surpresa (SCHERER, 2003; LAUKKA, 2004).

Na literatura não se encontra homogeneidade nos achados acústicos para a raiva, salvo a intensidade vocal que prevalece forte. Dessa forma, optou-se, nesse estudo, por analisar a entonação da fala com raiva comparada a fala neutra. Williams e Stevens (1972) encontraram, na raiva, contornos de F_0 superior às demais emoções. Laukka (2004) e Stibbard (2001) também encontraram altos valores de F_0 com grande variação a longo termo e contorno entonacional crescente. Em contraposição, alguns estudos encontraram frequência fundamental baixa (SOBIN; ALPERT, 1999), o que mostra incongruências entre os resultados. Como Mozziconacci (2002) sugere que as emoções sejam comparadas com a fala neutra do próprio sujeito, a fim de possibilitar comparações dentro e entre estudos. Nesse estudo, foi comparada a fala neutra com a fala com raiva. Assumimos a hipótese de que a voz neutra possui valores de F_0 sempre mais baixos, ou em um registro mais baixo do que aquelas vozes entoadas com raiva.

Metodologia

Participaram da pesquisa três atrizes profissionais com idades entre 40 e 50 anos, com média de 20 anos de profissão cada uma. A coleta dos dados se deu no laboratório do Estúdio Múltiplos do Centro de Computação Eletrônica da Universidade de São Paulo (CCE-USP). As atrizes foram instruídas a ler um texto informativo com cerca de 126 palavras, cujo tema era imunologia (VAZ, 1983), de modo que o léxico não viesse a interferir na interpretação. Cada uma das atrizes leu o texto de maneira atuada seguindo quatro emoções: raiva, alegria, medo e tristeza. Foi ainda previsto um grupo de controle chamado “fala neutra” em que as atrizes liam o texto sem inflexões emotivas. Dessa maneira, obteve-se um total de 75 gravações (4 emoções + 1 controle x 3 atrizes x 5 repetições). As gravações foram captadas com um mesa de som *Mackie CFX16*, com uma taxa de amostragem de 44KHz e uma resolução de 32 bit. Para o presente trabalho foi selecionada a emoção “raiva”, a qual será comparada com o grupo de controle “fala neutra”. Foi realizada a segmentação do texto em unidades menores (frases), selecionando uma frase para a análise. A segmentação linguística foi feita com base em Nespor e Vogel (1986) e descrita em pormenores em Vassoler (2011).

Foram usadas a primeira, a terceira e a quinta repetição de cada sujeito, obtendo-se um total de 18 tokens para análise (3 repetições x 2 emoções x 3 sujeitos). A unidade analisada foi:

- (1) *as células do sangue que fabricam anticorpos são individualizadas.*

A análise das variações de F_0 foram feitas usando a rotina de análise da prosódia ExProsodia (FERREIRA NETTO, 2010). Nessa análise, foi usado o método de detecção automática das sílabas, chamado de método *c.p.s.* (candidato a pico silábico) pela média.

O método mapeia valores de frequência fundamental e intensidade a cada 5 milissegundos, a fim de determinar se há ou não uma unidade entoacional naquele ponto. Para se obter uma unidade entoacional os valores de frequência e intensidade tem de se manter dentro dos limites escolhidos pelo analista e maiores que zero por mais de 20 milissegundos. Sendo satisfeitos todos estes parâmetros o programa acusará que naquele ponto há uma unidade entoacional. A consistência dos valores de F_0 é calculada usando-se testes de hipótese, como um *teste-z*, a fim de que não haja variações muito grandes no interior de uma unidade. Neste método, portanto, temos dois tipos de análise: a primeira mapeia os valores de F_0 (a unidades entoacionais), em busca de valores válidos, ao passo que a segunda analisa a constância de tais dados. O parâmetro de constância é sempre a média, valores muito distantes da média são excluídos.

No momento em que houver inconstâncias ou a frequência ou a intensidade igualam-se à zero ou ultrapassam os valores escolhidos pelo analista, a rotina compreenderá como o fim de uma unidade entoacional e recomeçará novamente o processo de mapeamento. Cada uma das orações foi analisada pelo método acima exposto. Os valores de média das unidades entoacionais das gravações foram calculados por meio da seguinte fórmula (1),

Onde:

τ = tom médio;

zx = unidade entoacional em mapeamento;

$zx + 1$ = unidade entoacional a ser mapeada.

A fórmula acima exposta nos diz que a média será obtida somando-se as unidades entoacionais válidas e tirando-se médias acumuladas até que a $zx + 1 = zx$, ou seja, haverá uma relação temporal intrínseca entre os momentos zx até que a última unidade mapeada pelo processo.

Análise

A partir dos valores obtidos foi feita uma média geral por unidade entoacional. Obtivemos, ao fim da análise, um total de 25 unidades para cada condição (raiva/fala neutra) em cada uma das repetições. Aquelas frases que possuíam menos do que 25 unidades entoacionais passaram por um processo de tratamento dos dados. Para este fim utilizamos um método de reposição de valores por amostragem, de modo que todos os vetores de dados pudessem ser equiparados, ou seja, ter a mesma quantidade de unidades entoacionais. Os valores obtidos em cada uma das unidades entoacionais foram somados aos seus pares (todas as primeiras unidade entoacionais, por exemplo, depois todas as segundas e assim por diante) e divididos por nove, o número de tokens por emoção. A partir desse procedimento, obteve-se uma média geral feita a partir da média das frequências das unidades entoacionais. Os valores apresentados no Quadro 1 resumem nossas estatísticas descritivas, as médias gerais foram de 197,54 Hz para a raiva, enquanto para a fala neutra obteve-se uma média de 156,13 Hz. Enquanto os valores de desvio-padrão geral foram de

7,46 para a primeira condição e 9,02 para a segunda. O coeficiente de variação intrínseco geral a cada condição foi de 3,78% e 5,78%, respectivamente. Os dados estão resumidos na tabela abaixo:

Tabela 1: Resumo dos valores das estatísticas descritivas

	Raiva	Neutra
Média	197,54	156,13
Desvio-padrão	7,46	9,02
Coef. de Variação	3,78%	5,78%

Os dois vetores de médias das unidades entoacionais de cada condição foram comparados por um teste de correlação de Pearson. O teste retornou um $\rho=0,88$, tal valor demonstra que ambas tem uma correlação muito forte, indicando que a despeito dos valores que as frequências fundamentais vieram a assumir em cada unidade entoacional a direção de ambos os vetores manteve-se a mesma. Foram também calculadas as correlações entre as condições entre sujeitos, tomando como base os valores de F_0 , de modo que todas foram maiores que 0,8 ($p>0,8$). Este valor indica que não se trata de uma idiosincrasia causada por conta das médias tomadas no processo de detecção automático. A Figura 1 mostra uma comparação entre as duas condições, os valores plotados no eixo y são as médias silábicas gerais, enquanto no eixo x se encontram as unidades entoacionais.

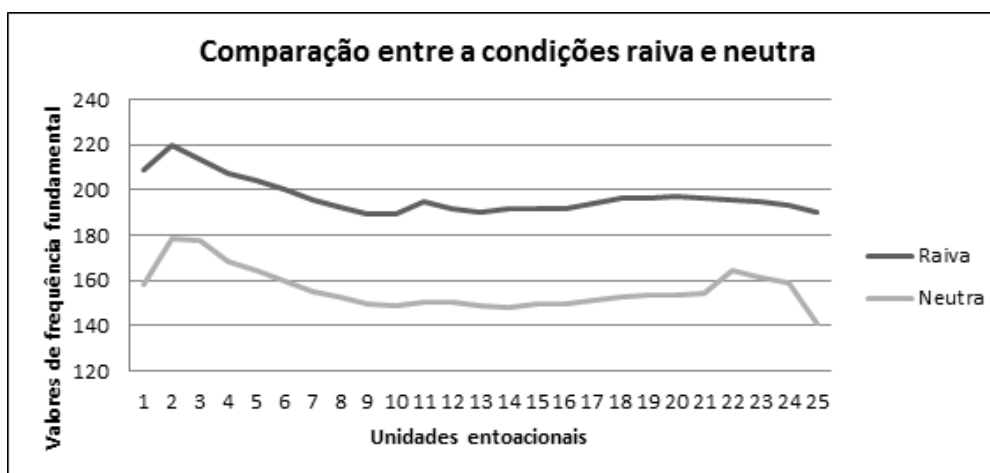


Figura 1: Comparação entre a fala neutra e a raiva, sendo que os valores plotados no eixo y correspondem as médias silábicas gerais, enquanto no eixo x se encontram as unidades entoacionais

Foram também calculados os coeficientes de variação para cada uma das condições. No caso da condição *raiva* obteve-se um coeficiente de variação de 3,78%, enquanto para a condição neutra houve um total de 5,78%. O que se pode notar é que a distribuição dos dados em ambos os casos é heterogênea, sugerindo que os dados estão alocados em intervalos bastante discriminados. Foi também executado um *teste-t* pareado, usando como fatores as condições raiva e neutra e como valores de análise as médias das unidades entoacionais, o primeiro teste foi executado usando-se um *bootstrapping*¹ com $n=10000$ e

¹ A técnica de *bootstrapping* consiste em reorganizar aleatoriamente, com base em um método de amostragem aleatória simples, um certo conjunto de dados. Em tese, o conjunto re-amostrado mantém as mesmas características de distribuição do conjunto de dados iniciais, uma vez que eles são amostras aleatórias de uma dada população.

$\alpha=5\%$ e um segundo sem *bootstrapping*, com $n=25$ e $\alpha=5\%$. Ambos os testes retornaram $p<0,001$, demonstrando que há, de fato, um efeito entre as duas amostras e que, provavelmente, ambas são diferentes.

Conclusão

Como se pode observar a condição “raiva” se comporta nas faixas de altas frequências, enquanto a frequência fundamental da fala neutra está localizada nas faixas de médias frequências, corroborando a literatura pesquisada. No entanto, o coeficiente de variação da raiva está em torno de 3,7 e 5,7 para a fala neutra, o que mostra mais consistência nas repetições da raiva. O desvio padrão da fala neutra (9) também foi maior se comparado com a raiva (7,4). No entanto, a correlação foi de 0,88 para ambas as situações, mostrando que a variação linear é quase idêntica.

Estes números apontam que, salvo variações intrínsecas ao sujeito analisado, as emoções possuem padrões de contorno muito específicos e bastante regulares. A diferença entre elas está nos *pitch ranges*. As frases entoadas com raiva tendem a ter um registro mais agudo, ao passo que as frases entoadas de maneira neutra mantêm-se em um registro ligeiramente mais central. A hipótese explicativa, de um ponto de vista fisiológico, refere-se à ideia de que a raiva atuada exige do ator um nível de enrijecimento dos músculos intercostais e das cartilagens laringais diferentes do que daqueles da voz neutra. Este aumento de tônus causaria um aumento na pressão pulmonar e, por consequência, do fluxo de ar e pressão subglotal, alterando os valores de frequência fundamental (TITZE, I. R.; SCHMIDT, S. S.; TITZE, M. R., 1995). Contudo, sem as medidas fisiológicas, tais hipóteses são apenas explicativas, de modo que a verificação demandaria uma nova pesquisa.

Do ponto de vista linguístico, é possível afirmar que há uma forma entoacional padronizada para a sentença escolhida, de modo que é a implementação fonética desta forma o fator determinante de seu valor modal. A figura 1 e o valor de correlação de Pearson demonstram que há certa regularidade neste sentido. A forma entoacional segue os princípios de teorias como a de Ladd (1996), as quais hipotetizam tal forma como um objeto mental, sendo que a realização segue as características do sujeito, bem como as condições de implementação (estado emocional, contexto conversacional e etc.). Ressaltamos que não podemos determinar com exatidão qual tipo de elocução é entoada com raiva ou neutra apenas com os dados acústicos, uma vez que a implementação é o fator determinante, como acima citado. Por outro lado, podemos vislumbrar a hipótese de que frases iguais entoadas de maneiras variadas mantêm certo padrão nas suas variações de frequência fundamental.

REFERÊNCIAS

- BANSE, R.; SCHERER, K. R. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, Genebra, v. 70, n. 3, p. 614-636, 1996.
- COLLIER, R.; ‘T HART, J. T.; COHEN, A. *A perceptual study of intonation*. Cambridge: Cambridge University Press, 1990.

DARWIN, C. R. *The expression of the emotions in man and animals*. London: John Murray, 1872.

FERREIRA NETTO, W. ExProsodia. *Revista da Propriedade Industrial – RPII*, Rio de Janeiro, v. 2038, n. 167, p. 167, 2010.

_____. Decomposição da entoação frasal em componentes estruturadoras e em componentes semântico-funcionais. In: *X Congresso Nacional de Fonética e Fonologia / IV Congresso Internacional de Fonética e Fonologia*. Niterói; UFF, 2008. p. 1-6.

_____. *Formação da prosódia da língua portuguesa*. 2006. Tese (Livre-Docência) – Universidade de São Paulo-USP, São Paulo.

HIRST, D.; DI CRISTO, A. A survey of intonation systems. In: HIRST, D.; DI CRISTO, A. (Ed.) *Intonation Systems*. Cambridge: Cambridge University Press, 1998.

JUN, S.-A. (Ed.) *Prosodic typology: the phonology of intonation and phrasing*. Oxford: Oxford University Press, 2006.

LADD, D. R. *Intonational phonology*. Cambridge: Cambridge University Press, 1996.

LADEFOGED, P. *Elements of acoustic phonetics*. Chicago & London: The University of Chicago Press, 1974.

LAUKKA, P. Vocal expression of emotion: Discrete-emotions and dimensional accounts. 2004. Tese (Doutorado). Uppsala, Sweden: Acta Universitatis Upsaliensis, 2004. v. 141. p. 1-80.

LIBERMANN, M. *The intonational system of English*. 1975. Tese (Doutorado em Linguística) - Massachusetts Institute of Technology, Cambridge, 1975.

MAEDA, S. *A Characterization of American English Intonation*. 1976. Tese (Doutorado em Engenharia Elétrica) – Massachusetts Institute of Technology – M.I.T., Cambridge.

MOZZICONACCI, S. J. L.; HERMES, D. J. Role of intonation patterns in conveying emotion in speech. *Proceedings of International Conference on Speech Prosody*. San Francisco: ISCA/SProSIG 1999. p. 2001-2004.

MOZZICONACCI, S. J. L. Prosody and Emotions. In: BEL, B.; MARLIEN, I. (Org.) *Proceedings of 1st International Conference on Speech Prosody*. Aix-en-Provence: ISCA/SProSIG, 2002. p. 1-9.

_____. Emotion and attitude conveyed in speech by means of prosody. In: *For the 2nd Workshop on Attitude, Personality and Emotions in User-Adapted Interaction*. Sonthofen, Germany, July, 2001, p. 1-10.

MURRAY, I. R.; ARNOTT, J. L.: Towards the Simulation of Emotion in Synthetic Speech: A Review of the Literature of Human Vocal Emotion. *Journal of Acoustic Society of America*, New York, v. 93, n. 2, p. 1097-1198, 1993.

- NESPOR, M.; VOGEL, I. *Prosodic Phonology*. Dordrecht: Foris Publications, 1986.
- OLIVEIRA, A. M. Aspectos acústicos da voz na representação teatral das emoções: um estudo preliminar. *Congresso Brasileiro de Fonoaudiologia*, 2008, Campos do Jordão, Sociedade Brasileira de Fonoaudiologia, p. 1312.
- OLIVEIRA, A. M.; MEDEIROS, B. R. Aspectos prosódicos de quatro emoções na voz falada. *VI Simpósio de Cognição e Artes Musicais*, Rio de Janeiro: ABCOGMUS, 2010. p. 292-300.
- PIKE, K. N. *The intonation of American English*. Michigan: Ann Arbor – University of Michigan Press, 1945.
- PIERREHUMBERT, J. *The Phonology of Phonetic of English Intonation*. 1980. Tese (Doutorado em Linguística) – Massachusetts Institute of Technology – M.I.T., Cambridge, 1980.
- SCHERER, K. R. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, v. 40, n. 1-2, p. 227-256, 2003.
- _____. Vocal measurement of emotion. In: PLUTCHIK, R.; KELLERMAN, H. (Org.). *Emotion: Theory, research, and experience. The measurement of emotion*. v. 4. New York: Academic Press, 1989. p. 233-260.
- _____ et al. On the nature and function of emotion: a component process approach. In: SCHERER, K. R.; EKMAN, P. (Org.) *Approaches to emotion*. Hillsdale, NJ: Lawrence Erlbaum, 1984. p. 293-318.
- SILVERMAN, K. et al. ToBI: a standard for labeling English Prosody *Proceedings, Second International Conference on Spoken Language Processing*, Alberta: ISCA. 1992, 2, 867-870.
- SOBIN, C.; ALPERT, M. Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research*, New York, v. 28, n. 4, p. 347-365, 1999.
- STIBBARD, R. M. *Vocal Expression of Emotions in Non-laboratory Speech: An Investigation of the Reading/Leeds Emotion in Speech Project Annotation Data*. Unpublished PhD thesis. University of Reading, Reading, UK, 2001.
- T'HART, J. On the anatomy of intonation. *Lingua*, Leiden, v. 19, n. 2, p. 177-19, 1967.
- _____. Differential sensitivity to pitch distance, particularly in speech. *Journal of Acoustical Society of the America*, New York, v. 69, n. 3, p. 811-821, 1981.
- T'HART, J.; COHEN, A. Intonation by rule: a perceptual quest. *Journal of Phonetics*, Bloomington, v. 1, n. 4. p. 309-327, 1973.
- TITZE, I. R.; SCHMIDT, S. S.; TITZE, M. R. Phonation threshold pressure in a physical model of the vocal fold mucosa. *Journal of Acoustical Society of the America*, New York, v. 97, n. 5, p. 3080-4, 1995.

VAISSIÈRE, J. Perception of intonation. In: PISONI, D. B.; REMEZ, R. E. (Ed.) *Handbook of Speech Perception*, Blackwell, 2004. p. 236-263.

VASSOLER, A. M. O. *O papel da frequência fundamental da fala na representação teatral das emoções*. 2011. Dissertação (Mestrado em Semiótica e Linguística Geral do Departamento de Linguística da Faculdade de Filosofia, Letras e Ciências Humanas) – Universidade de São Paulo – USP.

VAZ, N. M. Reconhecer a si próprio – Ideias para uma nova imunologia. *Ciência Hoje II*, v. 7, p. 32-8, 1983.

XU, Y.; WANG, Q. E. Pitch targets and their realization: Evidence from Mandarin *Chinese Speech Communication*, v. 33, n. 2, 319-337, 2001.

WILLIAMS, C. E.; STEVENS, K. N. Emotions and speech: Some acoustical correlates. *Journal of Acoustics Society of America*, New York, v. 52, p. 1238-1250, 1972.