

■ Análise automática de manifestações emocionais de tristeza e cólera em PB: abordagem pelo programa ExProsodia

WALDEMAR FERREIRA NETTO

Doutor em Linguística. Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, Brasil.
wafnetto@usp.br

DANIEL OLIVEIRA PERES

Doutorando em Filologia e Língua Portuguesa. Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, Brasil.
doliveiraperes@gmail.com

MARCUS VINÍCIUS M. MARTINS

Doutorando em Filologia e Língua Portuguesa. Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, Brasil.
marcusvmmartins@gmail.com

RENATA CEZAR DE MORAES ROSA

Mestranda em Filologia e Língua Portuguesa. Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, Brasil.
rentcmr@gmail.com

MARESSA DE FREITAS VIEIRA

Doutora em Filologia e Língua Portuguesa. Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, Brasil.
maressafv@gmail.com

Resumo: Manifestações emocionais na linguagem são uma preocupação desde o início do séc. XX. Tomando princípios do programa ExProsodia de análise automática da entoação, cinco parâmetros para a análise da fala emotiva no português do Brasil foram estabelecidos. Dados de fala espontânea foram coletados na internet. O processamento foi feito com o aplicativo ExProsodia. Os resultados, testados pela comparação múltipla de Tukey e pela análise de cluster, apontaram para a identificação da manifestação de cólera, mas não para manifestação de gênero. O gênero foi identificado apenas em relação às falas neutras e às tristes, embora não se identificassem essas manifestações isoladamente.

Palavra-chave: entoação, prosódia, manifestações emocionais

Abstract: Emotional manifestations in language are analysed since the beginning of the century XX. This paper presents five parameters used for analysis of emotional speech in Brazilian Portuguese. The spontaneous speech data were collected in digital files on Internet. Automatic processing was done with the application ExProsodia. The results of this analysis, tested by Tukey's multiple comparison and verified by multivariate analysis of cluster's test, showed identification of male and female anger, but not of gender. There were gender identification in relation to the neutral and the sad, but it was not possible to identify these emotional expressions alone.

Key-Words: intonation, prosody, emotional speech

Introdução

A análise da manifestação das emoções associadas à fala tem sido objeto de especulação científica desde o início do século XX. Weaver (1924a; 1924b; 1924c) 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 sessenta, a entoação emocional foi retomada. (MARKEL, 1965; CONSTANZO et al., 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 (SCHERER et alii, 1984) concluíram que, embora a entoação seja responsável pela manifestação de emoções, seu reconhecimento associa-se a traços gramaticais específicos e que seria necessário distinguir traços linguísticos e paralinguísticos de F0. Em 1998 (BATLINER et al. 2011), o trabalho de Slaney e McRoberts (1998) com dados espontâneos de fala dirigida às crianças, inova utilizando dados de fala espontânea na análise automática. Recentemente, um grande número de pesquisas têm analisado a 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 que 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 a variação da frequência média e da dispersão, medida pelo coeficiente de variação da frequência média, são parâmetros 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.

Neste trabalho, pretende-se verificar se um conjunto de parâmetros estabelecido pelo programa ExProsodia de análise automática da entoação do Português do Brasil permite reconhecer manifestações emocionais produzidas pelos sujeitos falantes.

2. O aplicativo ExProsodia

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 e seus colegas (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 F_0 , que se repete nos momentos $Z(t)$ mensurados de F_0 . 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 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 série temporal se configura aditivamente como $Z(t)=S(t)+F(t)+E(t)$.

O gráfico abaixo mostra o ritmo tonal num sistema de coordenadas cartesianas em que o eixo horizontal corresponde a medições temporais em milissegundos (ms) e o eixo vertical corresponde a medições de frequências em Hz. As setas diagonais escuras representam a declinação desencadeada em cada momento pela supressão do esforço e as setas verticais escuras representam a sustentação desencadeada pela retomada da tensão inicial, momento a momento, formando o ritmo tonal. Os pontos marcados com o símbolo Z são os momentos nesse processo em que há duração, intensidade e frequência suficientes para que os interlocutores envolvidos no processo da fala possam estabelecer cada uma das unidades formadoras da entoação frasal.

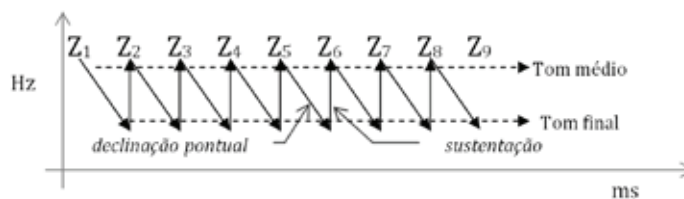


Fig. 1: Ritmo Tonal

O modelo de análise apresentado permite a análise isolada de cada uma das componentes de F0.

3. Unidade básica da entoação - UBI

A seleção das unidades $Z(t)$ — ora chamadas de UBI (*Unit of Base of Intonation*) — para análise é feita pelo aplicativo ExProsodia[®]. O aplicativo faz a análise automática de porções da curva de frequência estabelecida por autocorrelação pelo software Speech Filing System[®] (HUCKVALE, 2008; HUCKVALE et al., 1987). Três parâmetros são considerados 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.

A seleção de 50 Hz como frequência mínima deu-se pela manutenção de um intervalo de segurança. O mesmo fato ocorre para as frequências mais agudas, com limite em 700 Hz. Russo e Behlau (1993) verificaram que falantes masculinos do português brasileiro têm uma frequência fundamental em torno de 105 Hz, os do sexo feminino têm 213 Hz, crianças antes da puberdade, em média, 290 Hz e recém-nascidos, em torno de 440 Hz. Mortari (1990) encontrou para vozes infantis entre sete e nove anos uma variação entre 182 e 281 Hz, para meninos e meninas, sem diferenças marcantes entre eles. Andrade (2003) encontrou para vozes masculinas uma variação média oscilando entre 110 e 146,7 Hz e, para vozes femininas, uma concentração acentuada em torno de 203,5 Hz. Felipe e colegas (2006) propuseram a normatização entre 119 e 120 Hz, para vozes masculinas, e entre 206 e 207 Hz, para vozes femininas. Apesar da variação na frequência fundamental na fala de língua portuguesa do Brasil, é possível estabelecer algum limite que abarque essa variação. Russo (1999) propôs que a área da fala incluindo a frequência fundamental esteja desde 100 Hz até 8000 Hz, com intensidade variando entre 40 e 65 dB. Tendo em vista objetivar-se a análise automática da fala e a população ser bastante heterogênea, optou-se

por uma margem de erro mais extensa, sobretudo para as frequências mais graves, de maneira a não se rejeitar ocorrências falso-negativas. O programa ExProsodia de pesquisa considera uma margem de erro de 35%, a partir do valor médio mínimo de 100 Hz para vozes masculinas (RUSSO, 1999), para o qual se pressupõe um valor final de aproximadamente 67 Hz, ou uma quinta descendente (FERREIRA NETTO; CONSONI, 2008), e se estabelece um valor mínimo de 50 Hz. Para valores mais agudos, o limite foi estabelecido arbitrariamente em torno de 2 vezes o valor da frequência média máxima das vozes infantis (ANDRADE, 2003). Esses valores, mínimo, de 50 Hz e, máximo, de 700 Hz são os limites possíveis assumidos para a análise automática aceitar um momento Z_t como passível de ser UBI.

Tendo em vista especialmente as variações de frequência da onda sonora, a intensidade é tratada somente como parâmetro de avaliação de audibilidade dessa frequência. O aplicativo SFS analisa a variação de intensidade com uma taxa de amostragem de 200 Hz e apresenta os resultados como unidades de RMS (*root mean square*) dos picos de intensidade de porções de 25 ms. Ainda que variações de intensidade da onda sonora sejam reconhecidamente importantes, principalmente para as análises de ritmo, não foram tomadas como referência, por sujeitarem-se a fatores extralinguísticos, tais como possível variação de posição do falante em relação ao microfone. Um valor entre 1 e 10 estabelece um corte nos momentos que tenham de 1/1 a 1/10 do valor médio da intensidade, em RMS.

Boemio *et alii* (2005) verificaram que, embora ambos os hemisférios processem informações em duas velocidades específicas — 25-50 ms e 200-300 ms — no giro temporal superior, a conexão que se faz com o sulco temporal superior é enfatizada no hemisfério esquerdo na velocidade de 25-50 ms e no hemisfério

direito na velocidade de 200-300 ms. Dessa maneira ambos os hemisférios atuam nas tarefas de percepção linguística, mas cada um terá melhor especialização em tarefas específicas, no caso relativas à duração e à precisão da análise. Os valores apresentados estabelecem a possibilidade de ocorrerem resoluções diferenciadas e simultâneas em cada um dos hemisférios. Quanto aos valores mínimos entre 25-50 ms, outros autores encontraram resultados semelhantes (SCHAEFFER, 1966; HUGGINS, 1972; ROEDERER, 2008; STEVENS, 2000; HENRIQUE, 2002; MENEZES, 2003). Dada a necessidade de segmentar unidades de entoação na fala, a seleção de quatro momentos de análise do SFS, correspondendo a 5 ms cada um, estabelece uma duração mínima de 20 ms. Apesar de os valores mínimos propostos pelos autores já referidos estarem acima dessa opção, é seguro manter uma margem de erro maior para não se incorrer em falsos negativos. A finalização das porções significativas é decorre de um teste z para cada elemento do conjunto proposto como UBI.

4. Parâmetros estabelecidos

Com o propósito de se descrever automaticamente as condições de produção da fala pelo programa ExProsodia, foram definidos 13 parâmetros, considerando-se frequência, duração e intensidade:

Frequência	Duração	Intensidade
F0dp_TM	entreUBI_dp	IM_cv
F0_skew	intra_UBI_dp	IM_skew
mUBI	entreUBI_cv	
TM_mUBI	intra_UBI_cv	
TM_skew		
F/Eneg_cv		
F/Epos_cv		

Parâmetros relativos à frequência partem do TM definido pela análise do ExProsodia® para cada UBI, medido em Hz. *F0dp_TM* é o coeficiente de variação de F0; *F0_skew* é a simetria de F0; *mUBI* é o menor valor válido de UBI; *TM_mUBI* é a diferença entre o TM e o menor valor válido de UBI; *TM_skew* é a assimetria de TM; *F/Eposg_cv* é o coeficiente de variação dos valores da componente Foco/Ênfase que ocorreram acima do limite superior do TM; *F/Eneg_cv* é o coeficiente de variação dos valores de Foco/Ênfase que ocorreram abaixo do limite inferior do TM. Todos os valores foram definidos pela média aritmética dos dados analisados.

Parâmetros relativos à duração partem das definições de UBI realizadas pela análise do ExProsodia®, medidas em ms. O parâmetro *entreUBI_dp* é o desvio-padrão médio obtido a partir da finalização de uma UBI e o início da seguinte. Para essa definição, foi estabelecido o limite máximo de 500 ms. *intraUBI_dp* é o desvio-padrão médio da duração de uma UBI. *entreUBI_cv* e *intraUBI_cv* são, respectivamente, o coeficiente de variação da duração que ocorre entre as UBIs e nas UBIs.

Parâmetros relativos à intensidade partem das definições da intensidade média (IM) realizadas pela análise do ExProsodia®, medida em RMS para cada UBI. O parâmetro *IM_cv* é o coeficiente de variação médio da intensidade das UBIs. *IM_skew* é a assimetria média da intensidade verificada nas UBIs.

Tendo em vista que as gravações eram tanto de depoimentos quanto de entrevistas, não se consideraram parâmetros que envolvessem duração e quantidade de pausas.

5. Experimento

Para a realização do experimento consideraram-se dois conjuntos de variáveis: (a) gênero — masculino

e feminino — e (b) emoções — colérica, neutra e triste — combinadas entre si. Os dados foram coletados na internet em sites que disponibilizam podcasts.: “Webcombrasil” <<http://www.webcombrasil.com.br/>>, “A Voz do Brasil” <<http://www.ebcservicos.ebc.com.br/programas/a-voz-do-brasil>>, “Podcast UNESP” <http://podcast.unesp.br/>. Também foram coletadas gravações de vídeos no site Youtube <<http://www.youtube.com/>>. Todos os arquivos sonoros foram extraídos com o software Soundtap Streaming Audio Recorder® v2.11, produzido pela NCH Software Pty Ltd. Os arquivos sonoros foram segmentados com o programa Adobe Audition 3.0.1 build 8347.0 [Audition\3.0] Copyright© 1992-2008 Adobe Systems Incorporated and its Licensors. A análise e conversão da curva de frequência fundamental e da curva de intensidade para arquivos de texto foi realizada pelo software Speech Filing System Release 4.8/Windows SFSWin Version 1.7, produzido e registrado por Mark Huckvale, em 2010. Todas as demais análises foram feitas pelo aplicativo ExProsodia® registrado pela Universidade de São Paulo produzido por Waldemar Ferreira Netto (FERREIRA NETTO, 2008). As análises estatísticas foram realizadas pelo software KyPlot® version 2.0 beta 15 (32 bit) produzido e registrado por Kuichi Yoshida.

Para a análise das variáveis gênero e manifestação emocional, foram selecionados 60 arquivos sonoros, distribuídos em grupos de 10. Os grupos compreendiam: fala masculina colérica (HC), fala feminina colérica (MC), fala masculina neutra (HN), fala feminina neutra (MN), fala masculina triste (HT) e fala feminina triste (MT). A avaliação das falas como colérica, neutra e triste foi feita coletivamente pelos membros da equipe, com base nas gravações originais. As falas tristes referem-se todas a mortes de parentes próximos ou descrição de situações extremamente vexaminosas. As falas neutras foram tomadas de entrevistas nas quais o entrevistado, objeto

da coleta, tratava de assuntos relativos a sua profissão. As falas coléricas referem-se a situações de estresse ou de ameaças. Não houve restrições quanto à qualidade da gravação.

Para as análises quantitativas realizaram-se testes de ANOVA, com nível de significância de 0,05. Também se realizou teste de *Tukey*, para comparações múltiplas, com o mesmo nível de significância. Os resultados obtidos nas análises dos 13 parâmetros foram os que vão descritos na Tab. 1. Na tabela é possível verificar os valores de *p* obtidos nos testes de ANOVA. Todos os testes de ANOVA, em que $p < 0,01$, também apresentaram pelo menos um valor de $p < 0,05$, no teste de *Tukey* de comparações múltiplas.

frequência		duração		intensidade	
<i>F0dp_TM</i>	ANOVA $p < 0,01$	<i>entreUBI_dp</i>	NS	<i>IM_cv</i>	NS
<i>F0_skew</i>	ANOVA $p < 0,01$	<i>intra_UBI_dp</i>	ANOVA $p < 0,01$	<i>IM_skew</i>	NS
<i>mUBI</i>	NS	<i>entreUBI_cv</i>	NS		
<i>TM_mUBI</i>	ANOVA $p < 0,01$	<i>intra_UBI_cv</i>	NS		
<i>TM_skew</i>	NS				
<i>F/Eneg_cv</i>	NS				
<i>F/Epos_cv</i>	ANOVA $p < 0,01$				

Tab. 1: parâmetros analisados.

6. Tabulação e análise

Os valores médios obtidos para cada um dos parâmetros relativos às variáveis vão descritos na tabela 2. Somente os parâmetros que obtiveram algum resultado significativo vão dispostos nessa tabela. Para uma melhor compreensão da seleção feita, pode-se observar esses resultados na Tab. 2. Os gráficos que vão a seguir (Fig. 2 até Fig. 6) foram feitos a partir dessa Tab.2. Em todos eles, as linhas contínuas à esquerda representam a variação entre as manifestações emocionais. As linhas à direita representam a variação das manifestações emocionais entre os gêneros.

	MC	MN	MT	HC	HN	HT
<i>F0_skew</i>	-0,25	0,32	0,82	0,61	2,55	2,37
<i>TM_mUBI</i>	202,30	139,40	123,20	135,30	76,00	68,50
<i>intra_UBI_dp</i>	57,32	64,77	65,28	49,94	64,74	67,64
<i>F/Epos_cv</i>	0,78	1,02	1,02	0,93	1,19	1,23
<i>F0dp/TM</i>	89,18	55,65	60,69	72,21	46,07	40,06

Tab. 2: na tabela vão os parâmetros significativos dispostos nas linha e as variáveis comparadas, segundo gênero e tipo de manifestação emocional, nas colunas. Para a descrição de cada um, ver o texto.

F0dp/TM — Em relação ao parâmetro *F0dp/TM*, o teste ANOVA apontou uma diferença significativa no conjunto dos dados ($F_o=16.45$ e $p<0,001$). O teste de *Tukey*, por sua vez, no que diz respeito à relação entre a dispersão das frequência e o valor médio acumulado dessas frequências, apontou diferenças significativas. Na comparação entre manifestações emocionais, MC diferiu de MN ($t_o=-5.36$ e $p<0,001$) e de MT ($t_o=-4.55$ e $p<0,001$), nas falas femininas, e HC diferiu de HN ($t_o=-4.18$ e $p<0,01$) e de HT ($t_o=-5.14$ e $p<0,001$), nas masculinas. Não houve, entretanto, nenhuma variação significativa apontada pelo mesmo teste entre MN e MT, nem entre HN e HT. Na comparação entre falas de gêneros diferentes, apenas HT e MT apresentaram diferenças significativas ($t_o=3.30$ e $p<0,05$).

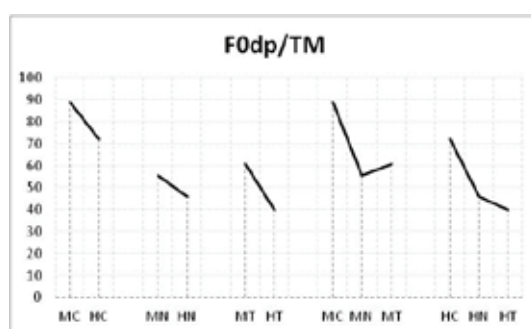


Fig. 2: Variação dos valores médios encontrados para o parâmetro *f0dp/tm*.

FO_skew — Em relação ao parâmetro *FO_skew*, o teste ANOVA apontou uma diferença significativa no conjunto dos dados ($F_0=8.81$ e $p<0,001$). O teste de *Tukey*, no que diz respeito à relação entre a assimetria da dispersão das frequência, apontou diferenças significativas. Na comparação entre manifestações emocionais, não houve diferença significativa para assimetria nas falas femininas. Nas falas masculinas, HC diferiu de HN ($t_0=3.58$ e $p<0,01$) e de HT ($t_0=3.24$ e $p<0,05$). Não houve diferença significativa apontada pelo mesmo teste, em relação a HN e HT. Na comparação entre os gêneros, MN e HN diferiram significativamente ($t_0=4.43$ e $p<0,01$).

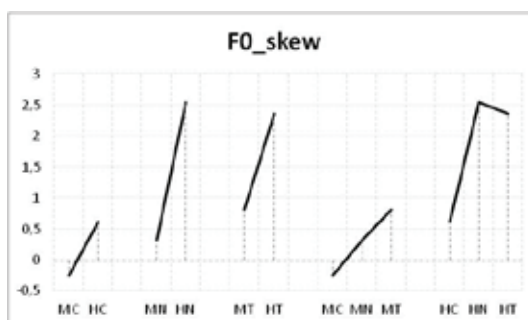


Fig. 3: Variação dos valores médios encontrados para o parâmetro *fo_skew*.

TM_mUBI — Em relação ao parâmetro *TM_mUBI*, o teste ANOVA apontou uma diferença significativa no conjunto dos dados ($F_0=14.68$ e $p<0,001$). O teste de *Tukey*, no que diz respeito à diferença entre o TM e a média da menor frequência válida encontrada, apontou diferenças significativas. Na comparação entre manifestações emocionais, MC diferiu de MN ($t_0=-3.50$ e $p<0,05$) e de MT ($t_0=-4.40$ e $p<0,001$), entre as mulheres, e HC diferiu de HN ($t_0=-3.30$ e $p<0,05$) e de HT ($t_0=3.72$ e $p<0,01$), entre os homens. Não houve, entretanto, diferença significativa entre MN e MT nem entre HN e HT. Na comparação entre os gêneros, houve diferença significativa em todas as comparações: entre MC e HC

($t_o=3.73$ e $p<0,01$), entre MN e HN ($t_o=3.52$ e $p<0,05$) e entre MT e HT ($t_o=3.04$ e $p<0,05$).

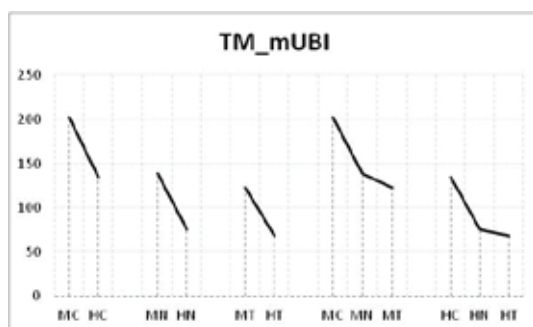


Fig. 4: Variação dos valores médios encontrados para o parâmetro *TM_mUBI*.

intraUBI_dp — Em relação ao parâmetro *intraUBI_dp*, o teste ANOVA apontou uma diferença significativa no conjunto dos dados ($F_o=6.40$ e $p<0,001$). O teste de *Tukey*, no que diz respeito à dispersão das durações das UBIs, apontou diferenças significativas. Na comparação entre as manifestações emocionais, não houve diferença significativa entre as falas femininas. Houve diferença entre HC e HN ($t_o=3.40$ e $p<0,01$) e entre HT ($t_o=4.73$ e $p<0,001$). Não houve diferença significativa entre HN e HT. Na comparação entre os gêneros não houve diferenças significativas.

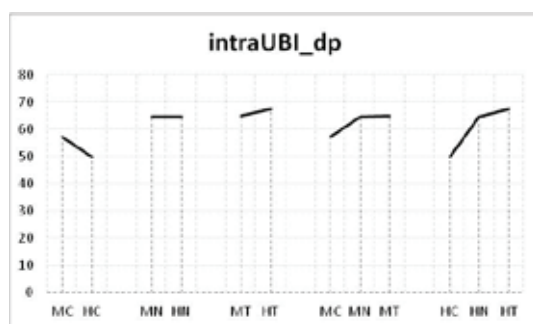


Fig. 5: Variação dos valores médios encontrados para o parâmetro *intraUBI_dp*.

F/Epos_cv — Em relação ao parâmetro *F/Epos_cv*, o teste ANOVA apontou uma diferença significativa no conjunto dos dados ($F_0=5.64$ e $p<0,001$). O teste de *Tukey*, no que diz respeito à relação entre a dispersão dos valores médios que ocorreram acima e abaixo do TM e essa mesma variação média, apontou diferenças significativas. Na comparação entre as manifestações emocionais, não houve diferenças significativas nas falas femininas. Nas falas masculinas, houve diferença entre HC e HT ($t_0=3.04$ e $p<0,05$). Não houve diferença entre HC e HN. Na comparação entre gêneros, não houve diferenças significativas.

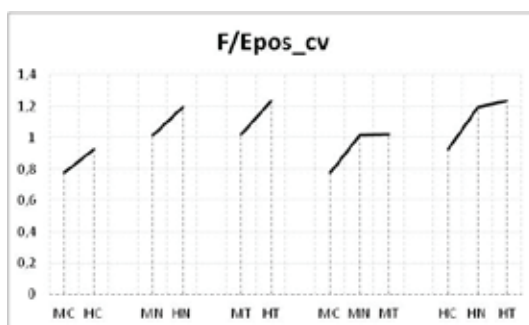


Fig 6: Variação dos valores médios encontrados para o parâmetro *F/Epos_cv*.

Com o propósito de se testar os parâmetros descritos acima, uma análise multivariada de Cluster pelo método de Ward (RUSSO et al., 2011) foi realizada, considerando-se um nível de significância de 0,05. Conforme se pode observar no dendrograma (Figura 7) resultante dessa análise, houve a identificação da manifestação de cólera, feminina e masculina, mas não a manifestação de gênero. Houve também a identificação de gênero em relação às falas neutras e às tristes, mas, nesse caso, não foi possível identificar nenhuma dessas emoções isoladamente.

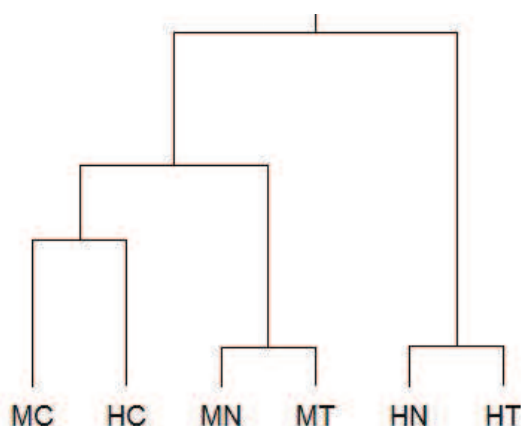


Fig. 7: no dendrograma acima, realizado pelo método de Ward (*Incremental Sum of Squares Method*), usando os dados da Tab. 2, é possível verificar no eixo horizontal as variáveis de gênero e de manifestação emocional agrupadas segundo os parâmetros estabelecidos. O eixo vertical representa a distância entre as variáveis agrupadas que se apontou por esse método: quanto menor o traço vertical maior a semelhança entre variáveis agrupadas ou grupo de variáveis.

7. Considerações Finais

As análises realizadas mostraram que os parâmetros apresentados foram suficientes para isolar a manifestação emocional de cólera das manifestações emocionais neutras. As manifestações neutras e tristes apenas puderam diferir-se quanto à variação de gênero, deixando isolada a variável que considerou o gênero masculino. Esse isolamento parece decorrer do TM que se caracterizaria pelos valores mais graves da fala masculina. De maneira semelhante, a discriminação da manifestação emocional de cólera parece decorrer da dispersão menor dos parâmetros analisados associada a um aumento significativo do TM.

Referências

- ANDRADE, L. M. O. *Determinação dos limiares de normalidade dos parâmetros acústicos da voz*. Dissertação (Mestrado) – USP, São Paulo, 2003.
- BÄNZINGER, T.; SCHERER, K. R. The role of intonation in emotional expressions. *Speech Communication*, n. 46, p. 252-267, 2005.
- BATLINER, A. et al. 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.
- BARBOSA, P. A. *Incursões em torno do ritmo da fala*. Campinas: Pontes; Fapesp, 2006.
- BAZ, D. G. B. *As relações entre entoação frasal e melodia de músicas populares paraguaias*. Tese (Doutoramento) – USP, São Paulo, 2011.
- BOEMIO, A. et al. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, v. 8, n. 3, p. 389-395, 2005.
- CONSONI, F. *Aspectos da percepção da proeminência tonal em português brasileiro*. Tese (Doutorado) – USP, São Paulo, 2011.
- COSTA, N. S. A. *Variações entoacionais na língua portuguesa falada por mulheres guatós*. Tese (Doutoramento) – USP/, FFLCH, São Paulo, 2010.
- COSTANZO, F. S.; MARKEL, N. N.; COSTANZO, P. R. Voice quality profile and perceived emotion. *Journal of Counseling Psychology*, v. 16, n. 3, p. 267-270, 1969.
- 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. 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.

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

FELIPPE, A. C. N.; GRILLO, M. H. M. N.; GRECHI, T. H. Normatização de medidas acústicas para vozes normais. *Revista Brasileira de Odontologia*, v. 72, n. 5, p. 659-64, 2006.

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, X/IV. Niterói, 2008.

FERREIRA NETTO, W. ExProsodia. *Revista da Propriedade Industrial* – RPI, 2038, pág. 167, item 120, em 26/out/2010.

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

FERREIRA NETTO, W.; CONSONI, F. Estratégias prosódicas da leitura em voz alta e da fala espontânea. *Alfa*, n. 52, v. 2, p. 521-534, 2008.

FERREIRA NETTO, W.; CONSONI, F.; PERES, D. O. Finalizações de frase em leituras e fala espontânea no PB. In: SEMINÁRIO DO GEL, 57. Ribeirão Preto, 2009.

MARTINS, M.; FERREIRA NETTO, W. Prosódia e escalas de frequência: um estudo em torno da escala de semitons. *Revel*, v. 8, n. 15, 2010.

HENRIQUE, L. L. *Acústica musical*. Lisboa: Calouste Gulbenkian, 2002.

HUCKVALE, M. A. *Speech Filing System v.4.7/Windows SFSWin Version 1.7*, em 17/02/2008.

HUCKVALE, M. A. et al. The SPAR Speech Filing System, In: EUROPEAN CONFERENCE ON SPEECH TECHNOLOGY, Edinburgh, 1987.

HUGGINS, A. W. F. Just noticeable differences for segment duration in natural speech. *Journal of Acoustical Society of America*, v. 51, n. 4., p. 1970-8, 1972.

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

MARTINS, M.; FERREIRA NETTO, W. Speech intonation and perception: a study of frequency scales for Brazilian Portuguese. *Journal of Acoustical Society of the America*, v. 129, n. 4, pt. 2, abr. 2011.

MENEZES, F. *A acústica musical em palavras e sons*. São Paulo: Ateliê Editorial; Fapesp, 2003.

MORETTIN, P. A.; TOLOI, C. M. *Séries temporais*. São Paulo: Atual, 1986.

MORTARI, A. L. *Análise instrumental da frequência fundamental e da intensidade da voz de crianças e adolescentes*. Dissertação (Mestrado) – PUC, São Paulo, 1990.

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

PEREIRA, B. B.; PAIS, M. B. Z.; SALES, P. R. H. *Análise espectral de séries temporais*. Rio de Janeiro: Arte Final leasing Editoria/Eletróbrás, 1986.

PERES, D.; CONSONI, F; FERREIRA NETTO, W. A influência da cadeia segmental na percepção de variações tonais. *LLJournal*, v. 6, n. 1, 2011.

PERES, D. O. *O papel da prosódia na identificação das variedades regionais do português brasileiro*. Dissertação (Mestrado) – USP, São Paulo, 2011.

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.

ROEDERER, J. G. *The physics and psychophysics of music*. An Introduction. 4th ed. New York: Springer, 2008.

RUSSO, I.; BEHLAU, M. *Percepção da fala: análise acústica do português brasileiro*. São Paulo: Lovise, 1993.

RUSSO, I. *Acústica e psicoacústica aplicadas à fonoaudiologia*. São Paulo: Lovise, 1999.

RUSSO, S. L.; CAMARGO, M. E.; SIMON, V. H.
Multivariate analysis of sonic synthetic profile in well. *Australian Journal of Basic and Applied Sciences*, v. 5, n. 8, p. 1090-1095, 2011.

SCHAEFFER, P. *Traité des objets musicaux* — Essai Interdisciplines. Paris: Éditions du Seuil, 1966.

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

SCHERER, K. R.; LADD, D. R; SILVERMAN, E. A. Vocal cues to speaker affect: testing two models. *Journal of Acoustical Society of Americas*, v. 76, n. 5, p. 1348-1356, 1984.

SKINNER, E. R. A calibrated recording an analysis of the pitch, force an quality of vocal tones expressing happiness an sadness. *Speech Monographs*, v. 2, n. 2, p. 81-137, 1935.

SLANEY, M.; McROBERTS, G. Baby ears: a recognition system for affective vocalization. In: *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*. Volume II: Features for automatic speech recognition. Seattle: The Institute of Electrical and Electronic Engineers Signal Processing Society, 1998. p. 985-988.

SPIEGEL, M. *Estatística*. São Paulo: McGraw-Hill, 1985.

STEVENS, K. *Acoustic Phonetics*. Cambridge: The MIT Press, 2000.

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.

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. 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 Journal of Applied Psychology*, v. 8, n. 1, p. 23-51, 1924b.

WEAVER, A. T. Experimental studies in vocal expression. *The Journal of Applied Psychology*, v. 8, n. 2, p. 159-186, 1924c.

WEAVER, A. T. Experimental studies in vocal expression. *The Quarterly Journal of Speech Education*, v. 10, n. 3, p. 199-204, 1924a.

WILLIAMS, C. E.; STEVENS, K. N. Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America*, v. 52, n. 4, p. 1238-1250, 1972.

XU, Yi; WANG, Q. E. Component of intonation: what are linguistic, what are mechanical/physiological?
In: INTERNATIONAL CONFERENCE ON VOICE PHYSIOLOGY AND BIOMECHANICS, Evanston Illinois, 1997.

[Recebido em 26 de junho de 2013
e aceito para publicação em 22 de novembro de 2013]