

# Percepção intercultural e interlinguística de emoções autênticas: Estudo com sujeitos brasileiros e suecos

Wellington da Silva<sup>1</sup>, Plínio A. Barbosa<sup>1</sup>

<sup>1</sup> Instituto de Estudos da Linguagem, Universidade Estadual de Campinas

wellingtondasilva@rocketmail.com, pabarbosa.unicampbr@gmail.com

## Resumo

Com o objetivo de investigar se a cultura e a língua materna influenciam na percepção de emoções expressas pela fala, conduzimos dois experimentos de percepção com sujeitos brasileiros e suecos, nos quais enunciados com emoções autênticas em português brasileiro e em sueco foram apresentados. Os sujeitos avaliaram, em escalas de 5 pontos, o grau de expressão de emoções básicas descritas por oito adjetivos nos enunciados em português brasileiro e o grau de expressão de cinco dimensões emocionais nos enunciados em sueco. A técnica de PCA revelou que dois componentes explicam mais do que 94% da variância das respostas dos juízes nos dois experimentos. Esses componentes foram preditos através de regressões lineares múltiplas por doze parâmetros acústicos automaticamente extraídos dos enunciados. Os resultados mostraram que a percepção das emoções expressas nos dois corpora foi muito semelhante entre ambas as culturas e que os sujeitos de ambas as nacionalidades se guiaram por estes mesmos parâmetros acústicos para realizar os julgamentos nos dois experimentos: inclinação do LTAS, mediana e semi-amplitude entre quantis da frequência fundamental e média e desvio-padrão da inclinação espectral. Concluímos, portanto, que a cultura e a língua materna dos sujeitos brasileiros e suecos não influenciaram na percepção das emoções.

**Palavras-chave:** fala expressiva, emoções básicas, dimensões emocionais

## 1. Introdução

O ser humano possui a característica de ser social e, por isso, está em constante interação com seus coespecíficos. As emoções são fundamentais nesse processo e desempenham um papel importante na formação e na manutenção de laços sociais. A maioria dos estudiosos das emoções humanas se concentrou durante décadas nas expressões faciais e corpóreas resultantes dos eventos desencadeadores das emoções. Porém, a fala também é um importante meio de expressão dos afetos, tendo recebido muita atenção nas últimas décadas, em parte por causa do desenvolvimento de meios de comunicação como o rádio e o telefone e, mais recentemente, de sistemas de síntese de fala e de reconhecimento automático do falante [1]. A influência do estado emocional do falante na sua voz ocorre porque as emoções causam, entre outras reações, respostas fisiológicas no indivíduo. Essas alterações, por sua vez, ocasionam variações na respiração, fonação e articulação, processos diretamente relacionados à fala [2]. Além disso, as emoções também podem afetar a atenção e a cognição do falante, dificultando o planejamento da fala (no caso do medo

ou da ansiedade, por exemplo) ou facilitando-o (no caso da alegria ou felicidade, por exemplo) [3].

Charles Darwin, em seu livro de 1872, apresentou a hipótese de que as emoções são produtos da evolução e foram selecionadas por auxiliarem na sobrevivência de nossos ancestrais, preparando-os para enfrentar situações relevantes no ambiente por meio de alterações corporais [4]. Essa visão evolucionista das emoções implica a noção de universalidade: se as emoções são um construto da evolução, então as mesmas emoções devem se manifestar de forma semelhante em todas as culturas, uma vez que os seres humanos constituem uma única espécie. A universalidade é, portanto, condição *sine qua non* da teoria evolucionista das emoções [5]. Desde a publicação desse livro de Darwin, pesquisadores têm investigado de forma mais elaborada a hipótese da universalidade da expressão de algumas emoções. Essas pesquisas, realizadas em sua maior parte com expressões faciais, revelaram que essa universalidade existe somente até certo ponto, mas que diferenças na expressão e no reconhecimento de algumas emoções também existem entre culturas. Semelhanças e diferenças interculturais também são encontradas nos estudos de expressão de emoções pela fala [6].

Entre as perspectivas teóricas que buscam explicar o que são e quais as funções das emoções, duas se destacam mais. A primeira delas, conhecida como *discrete emotion theory*, entende que os processos emocionais são distintos entre si e possuem características (eventos antecedentes, avaliação, alterações fisiológicas, características expressivas e comportamentais) únicas e universais. Essas emoções são conhecidas na literatura como “básicas” e sua existência está limitada a um pequeno número que varia de 6 a 14, dependendo do pesquisador [7]. Elas recebem nomes das línguas naturais, tais como “alegria”, “raiva”, “tristeza”, “medo”, “nojo” etc. A outra perspectiva é a *visão dimensional das emoções*, para a qual todos os estados emocionais podem ser caracterizados com relação ao grau de algumas dimensões emocionais, tais como ativação, prazer intrínseco, dominância e outras. Essas dimensões se manifestam quanto à sua intensidade por um contínuo, e não apenas por pólos discretos de máxima ou mínima intensidade, como sugerido pela perspectiva das emoções básicas [8].

O principal objetivo do estudo apresentado foi o de investigar, dentro dessas duas perspectivas teóricas, se a cultura e a língua materna dos sujeitos influenciam na percepção das emoções, ou seja, se o fato de os estímulos serem de uma língua estrangeira (português brasileiro ou sueco) da qual os sujeitos não possuem nenhum conhecimento afeta a percepção das emoções pela fala. Também avaliamos

se os sujeitos brasileiros e suecos se guiam pelos mesmos parâmetros prosódicos para julgar a expressão das emoções.

## 2. Experimento I

### 2.1. Material e métodos

#### 2.1.1. Corpus

Os estímulos desse experimento consistiram de trechos de fala de depoimentos reais de mulheres falantes do português brasileiro, extraídos do filme nacional de estilo documentário “Jogo de cena”, de 2007. Foram selecionados 30 trechos de 8 mulheres, cada um com duração variando de 3 a 10 segundos e com qualidade boa o suficiente para garantir análises acústicas adequadas. O áudio de cada trecho foi extraído em formato *wave*, mono, a uma taxa de amostragem de 44.100 Hz.

#### 2.1.2. Sujeitos

**Brasileiros: 26 sujeitos** (9 homens e 17 mulheres) completaram esse experimento, todos tendo residido a maior parte da vida no Brasil, universitários (cursando graduação ou pós-graduação) e sem perda de audição. A média de idade desse grupo foi de 23 anos (variando de 18 a 35 anos).

**Suecos: 16 sujeitos** (7 homens e 9 mulheres), todos eles tendo vivido a maior parte da vida na Suécia e falantes nativos da língua sueca. Eles também reportaram não entender a língua portuguesa nem ter problemas auditivos. Todos eram estudantes de graduação ou de pós-graduação da Universidade de Gotemburgo, Suécia. A média de idade desses sujeitos foi de 26 anos (entre 20 e 47 anos).

#### 2.1.3. Procedimento

Nesse experimento os sujeitos julgaram, em escalas de 5 pontos graduadas de 0, “nada **ADJETIVO**”, a 4, “muito **ADJETIVO**”, o grau com o qual o falante de cada enunciado expressava as emoções discretas descritas por oito adjetivos (alegre, comovido, surpreso, triste, contente, angustiado, aflito e entusiasmado). O experimento foi composto de 8 partes, sendo que em cada uma delas os sujeitos julgaram um adjetivo diferente, para todos os 30 trechos de fala, que eram apresentados aleatoriamente. Para evitar o cansaço dos juízes, o experimento foi conduzido em duas sessões, realizadas em dias diferentes. Em cada sessão, os sujeitos avaliaram a emoção expressa por quatro itens lexicais diferentes. Na primeira, foram avaliadas as emoções *alegre*, *comovido*, *surpreso* e *triste* e na segunda, *contente*, *angustiado*, *aflito* e *entusiasmado*, exatamente nessa ordem.

O experimento foi elaborado e aplicado via internet por meio da plataforma “*Survey Gizmo*”, disponível em “<http://www.surveygizmo.com>”. Os textos do experimento (instruções, questionários e adjetivos) foram apresentados na língua materna dos sujeitos, ou seja, em português para os sujeitos brasileiros e em sueco para os sujeitos suecos. Era apresentado na tela um estímulo por vez junto com sua respectiva escala. O áudio era reproduzido automaticamente assim que a página do navegador de internet terminava de carregar. Os sujeitos foram instruídos a usar fone de ouvido e a fazer o experimento individualmente em um ambiente silencioso. As respostas dos juízes na escala de 0 a 4 foram convertidas em uma escala linear de 0 a 1 (0, 0,25, 0,50, 0,75 e 1), igualando, assim, a escala desses valores com aquela dos valores normalizados por *z-score* dos parâmetros acústicos.

## 2.2. Análises e resultados

### 2.2.1. Análise dos julgamentos perceptivos dos sujeitos

A concordância entre as respostas dos juízes foi verificada pelo cálculo do índice kappa de Fleiss para cada adjetivo. Devido ao limite de espaço essas análises foram omitidas do presente artigo, mas podem ser consultadas em [9]. Os índices kappa foram estatisticamente significativos para os oito adjetivos e apontaram concordância satisfatória entre os sujeitos de ambas as nacionalidades.

A fim de reduzir o número de dimensões para as análises seguintes e evitar variáveis correlacionadas entre si, foi realizada uma Análise de Componentes Principais (PCA) com a média das respostas dos juízes no teste de percepção em cada enunciado para os oito adjetivos. A PCA é um procedimento estatístico que tenta encontrar o menor número de combinações lineares das variáveis originais necessário para explicar a maior parte da variância dos dados. A análise mostrou que dois componentes principais são suficientes para explicar mais do que 90% da variância total das respostas dos dois países. Para os sujeitos brasileiros, esses dois componentes explicam juntos **96,2%** da variância, sendo que o primeiro (PC1) explica 80,8% da variância e o segundo (PC2), 15,4%. Para os sujeitos suecos, dois componentes explicam juntos **94,5%** da variância total das respostas. O PC1 explica 80,3% da variância e o PC2, 14,2%. A análise dos *loadings*, que correspondem ao peso de cada variável em um determinado componente, revela como os sujeitos julgaram os oito adjetivos (Fig. 1). Observa-se na Fig. 1 que o primeiro componente de ambas as nacionalidades separa as etiquetas emocionais em dois grupos, um positivo e outro negativo, que colocam as etiquetas alegre, contente e entusiasmado de um lado e comovido, triste, angustiado e aflito, do outro.

### 2.2.2. Análise fonético-acústica

Para determinar os parâmetros acústicos nos quais os juízes se basearam para inferir o estado emocional dos falantes, doze parâmetros acústicos foram extraídos automaticamente dos enunciados do corpus por um script implementado para o software Praat [10, 11]. Esses parâmetros abrangem as classes de frequência fundamental ( $f_0$ ), primeira derivada de  $f_0$  ( $df_0$ ), intensidade global, inclinação espectral e espectro médio de longo-termo (LTAS). Eles são os seguintes:

**$f_0$**  (normalizada por *z-score*): mediana, semi-amplitude entre os quantis 95% e 5%, quantil 99,5% e assimetria;

**$df_0$** : média, desvio-padrão e assimetria;

**intensidade global**: assimetria;

**inclinação espectral** (diferença de intensidade entre as bandas de 0 – 1250 Hz e 1250 – 4000 Hz): média, desvio-padrão e assimetria;

**LTAS**: inclinação (diferença de intensidade entre as bandas de 0 – 1000 Hz e 1000 – 4000 Hz).

Em seguida, os dois componentes da PCA dos sujeitos das duas culturas foram correlacionados com os parâmetros acústicos por meio de regressões lineares múltiplas. Por meio do coeficiente de determinação ( $R^2$ ) que essa análise fornece, pode-se avaliar o quanto a variância da combinação de parâmetros acústicos explica a variância dos *scores* dos componentes da PCA. Os valores de  $\beta$  correspondem aos coeficientes normalizados do modelo de regressão linear.

### 3. Experimento II

#### 3.1. Material e métodos

##### 3.1.1. Corpus

O corpus utilizado no experimento II é composto de 40 trechos de fala com conteúdo emocional de 5 falantes femininos do sueco, com duração variando de 1 a 6 segundos e com qualidade boa o suficiente para garantir análises acústicas adequadas. Os enunciados foram extraídos de situações reais (debates e entrevistas) de programas transmitidos por uma emissora de televisão sueca e por um programa de rádio sueco disponibilizado livremente na internet como *podcast*. O áudio dos trechos selecionados foi extraído em formato *wave*, mono, a uma taxa de amostragem de 44.100 Hz.

##### 3.1.2. Sujeitos

**Suecos: 19 sujeitos** (6 homens e 13 mulheres) completaram o experimento. Eles eram estudantes de graduação ou pós-graduação ou funcionários da Universidade de Gotemburgo, Suécia. A idade média foi de 30 anos, variando de 21 a 56 anos. Todos eles nasceram e viveram a maior parte da vida na Suécia e possuem o sueco como língua materna. Eles também informaram não ter problemas auditivos.

**Brasileiros: 20 sujeitos** completaram o experimento (7 homens e 13 mulheres), com média de idade de 25 anos (variando de 18 a 34 anos). Todos eram estudantes de graduação ou de pós-graduação e nasceram e viveram a maior parte da vida no Brasil. Todos os sujeitos informaram não ter nenhum conhecimento da língua sueca e nenhum problema auditivo.

##### 3.1.3. Procedimento

Nesse experimento os sujeitos julgaram, em escalas de 5 pontos graduadas de 0, “nada ADJETIVO”, a 4, “muito ADJETIVO”, o grau com o qual o falante de cada enunciado expressava o estado emocional descrito por *dimensões emocionais*. Este experimento consistiu de cinco partes, realizadas em sessão única. Em cada uma, os sujeitos ouvintes avaliaram uma dimensão emocional para todos os 40 trechos de fala. As dimensões emocionais avaliadas foram: *ativação* (“Quão agitado estava o falante?”), *justiça* (“Quão justo o falante considerou o que aconteceu?”), *valência* (“Quão agradável para o falante foi a situação na qual ele estava?”), *motivação* (“Quão motivado para agir na situação está o falante?”) e *envolvimento* (“Quão envolvido com a situação o falante está?”), nesta ordem.

Os trechos de fala eram apresentados aos sujeitos aleatoriamente e eles deveriam avaliar, para cada um, o grau de expressão da dimensão emocional específica àquela parte do experimento selecionando um valor na escala. O restante do procedimento é o mesmo que o seguido no experimento I.

#### 3.2. Análises e resultados

Assim como no experimento I, a concordância entre as respostas dos sujeitos foi verificada pelo cálculo do índice kappa de Fleiss para cada dimensão emocional. Todas as cinco dimensões tiveram um nível de concordância estatisticamente significativo e aceitável para as análises posteriores [9].

Uma Análise de Componentes Principais (PCA) foi realizada com as cinco dimensões emocionais, usando a média das respostas dos sujeitos das duas culturas para cada trecho

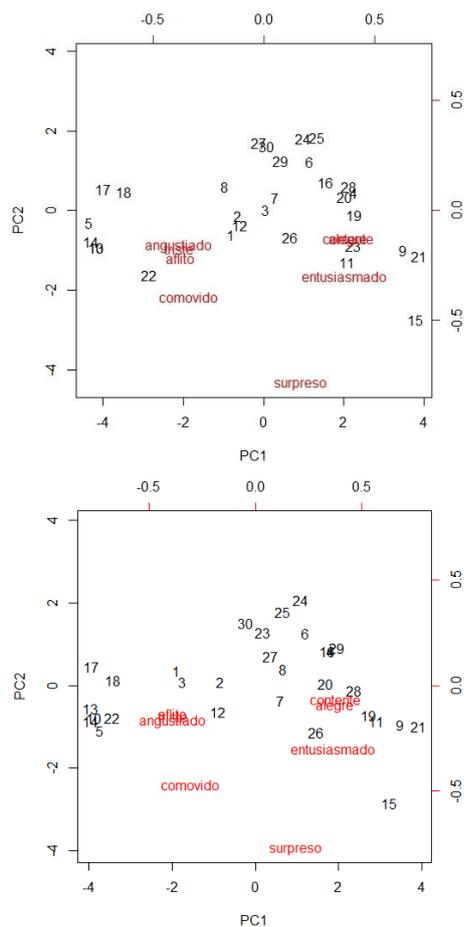


Figura 1: Adjetivos do experimento I projetados de acordo com seus loadings e os trechos de fala do corpus projetados em função de seus scores, para o PC1 e PC2. Superior: julgamentos dos sujeitos brasileiros. Inferior: julgamentos dos sujeitos suecos.

Para o PC1 dos sujeitos brasileiros, o melhor modelo de regressão foi o que combinou os parâmetros assimetria da  $f_0$  ( $\beta = 0,23$ ), assimetria da intensidade ( $\beta = 0,16$ ) e média ( $\beta = -0,60$ ) e desvio-padrão da inclinação espectral ( $\beta = -0,20$ ), que gerou um  $R^2$  ajustado de **34%** ( $F[4, 25] = 4,717, p < 0,006$ ). Para o PC1 dos sujeitos suecos, a combinação dos parâmetros assimetria da  $f_0$  ( $\beta = 0,21$ ), média ( $\beta = -0,46$ ) e desvio-padrão da inclinação espectral ( $\beta = -0,21$ ) explicou **31%** ( $F[3, 26] = 5,431, p < 0,005$ ) de sua variância.

O modelo que mais explicou a variância do PC2 dos sujeitos brasileiros foi o que combinou os parâmetros mediana da  $f_0$  ( $\beta = -0,58$ ) e assimetria da inclinação espectral ( $\beta = -0,29$ ), que gerou um  $R^2$  ajustado de **40%** ( $F[2, 27] = 10,46, p < 0,0005$ ), enquanto que a combinação dos parâmetros mediana da  $f_0$  ( $\beta = -0,39$ ), semi-amplitude entre quantis da  $f_0$  ( $\beta = -0,44$ ), e assimetria da inclinação espectral ( $\beta = -0,24$ ) explicou **53%** ( $F[3, 26] = 12,07, p < 0,00004$ ) da variância do PC2 dos sujeitos suecos.

Assim, essa análise mostra que, excetuando-se a assimetria da intensidade para o PC1 e a semi-amplitude entre quantis da  $f_0$  para o PC2, os parâmetros acústicos que explicam as respostas dos sujeitos de ambas as nacionalidades são os mesmos.

de fala. Para os sujeitos suecos, a análise revelou que dois componentes explicam **98,2%** da variância total das respostas. O primeiro desses componentes (PC1) explica 86% dessa variância, enquanto que o segundo (PC2), 12,2%. Para os sujeitos brasileiros, dois componentes explicam **95,4%** da variância total das respostas. O PC1 é responsável por 76% dessa variância, enquanto que o PC2, é responsável por 19,4% da variância. A análise dos *loadings* que as dimensões emocionais apresentaram para o PC1 mostra que os sujeitos de ambas as nacionalidades distinguiram nos enunciados um estado de “calma” (avaliado com valores maiores para as dimensões “valência” e “justiça”) de um estado de agitação emocional (avaliado pelos juízes com valores maiores para as dimensões “ativação”, “motivação” e “envolvimento”).

Os doze parâmetros acústicos descritos na seção 2.2.2 também foram extraídos automaticamente para os 40 trechos de fala usados como estímulos nesse experimento. A análise de regressão linear múltipla foi utilizada para associar esses parâmetros acústicos com os dois componentes da PCA dos sujeitos brasileiros e suecos. Para o PC1 dos sujeitos suecos, o melhor modelo combinou os parâmetros mediana ( $\beta = 0,36$ ) e semi-amplitude entre quantis da  $f_0$  ( $\beta = 0,05$ ), média ( $\beta = 0,18$ ) e desvio-padrão da inclinação espectral ( $\beta = 0,15$ ), e inclinação do LTAS ( $\beta = -0,83$ ), e explicou **79%** ( $F[5, 34] = 30,32, p < 10^{-10}$ ) da sua variância. Para o PC1 dos sujeitos brasileiros, a combinação dos parâmetros mediana da  $f_0$  ( $\beta = -0,43$ ) e inclinação do LTAS ( $\beta = 0,62$ ) explicou **82%** ( $F[2, 37] = 87,43, p < 10^{-14}$ ) da sua variância.

Não foi possível prever significativamente o PC2 das duas nacionalidades através dos parâmetros acústicos utilizados. Contudo, esse componente por si só não é tão importante quanto o PC1, visto que explica somente uma pequena proporção da variância das respostas dos sujeitos. Os parâmetros mediana da  $f_0$  e inclinação do LTAS, os quais foram os únicos que explicaram significativamente o PC1 dos sujeitos brasileiros, também foram os que apresentaram o maior peso no modelo de regressão com o PC1 dos sujeitos suecos (os maiores valores de  $\beta$ ).

#### 4. Discussão geral e conclusão

O presente estudo foi conduzido para testar experimentalmente a hipótese da universalidade da expressão de emoções pela fala, isto é, para investigar se a cultura e a língua materna influenciam na percepção de emoções expressas pela fala. Para isso, avaliou-se a percepção de sujeitos brasileiros e suecos de emoções expressas em trechos de fala na sua língua materna e também em uma língua estrangeira da qual os sujeitos não possuíam nenhum conhecimento (português brasileiro ou sueco).

As análises de componentes principais, que se mostraram muito robustas em organizar ortogonalmente os estados emocionais independentes que subjazem os julgamentos dos sujeitos, revelaram que os julgamentos perceptivos dos sujeitos de ambas as nacionalidades foram muito semelhantes. Além disso, as análises de regressão linear entre os parâmetros acústicos e os componentes da PCA mostraram que, de um modo geral, os parâmetros acústicos mais relevantes para explicar as respostas dos sujeitos nos dois experimentos são os mesmos para as duas nacionalidades. Os resultados foram condizentes com aqueles encontrados na literatura, segundo os quais um aumento no estado de agitação emocional do falante está relacionado a um aumento da frequência fundamental da voz e a um reforço maior da energia concentrada nos

harmônicos de frequências elevadas (refletido nos parâmetros inclinação do LTAS e inclinação espectral). É importante ressaltar que o LTAS e a inclinação espectral são parâmetros correlatos da qualidade de voz [12].

Outro resultado a ser observado é o fato de a frequência fundamental, que é usada de maneira diferente pelo português brasileiro e pelo sueco no plano linguístico, ter sido interpretada de forma igual pelos sujeitos brasileiros e suecos no julgamento da expressão das emoções e dimensões emocionais (plano extralinguístico).

É possível concluir, portanto, que a cultura, a língua materna e a experiência emocional dos sujeitos brasileiros e suecos não influenciaram na percepção das emoções expressas nos dois corpora. Isso sugere que a percepção da expressão de emoções básicas e de primitivos emocionais pela fala não depende da língua materna ou da cultura do ouvinte, pois as alterações nos parâmetros acústicos causadas pelas emoções seriam interpretadas universalmente de uma mesma maneira, como é o caso das expressões faciais das emoções. Entretanto, estudos com outras línguas de estruturas prosódicas diferentes entre si devem ser conduzidos para reforçar essa conclusão.

#### 5. Agradecimentos

O presente estudo foi conduzido com o apoio das bolsas de Mestrado (processo 2012/04254-4) e de Estágio de Pesquisa no Exterior (processo 2013/06082-9) concedidas pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) ao primeiro autor. Durante a redação deste artigo, o primeiro autor teve o apoio da bolsa de doutorado do CNPq.

#### 6. Referências

- [1] K. R. Scherer, “Vocal communication of emotion: a review of research paradigms”. In: *Speech Communication*, v. 40, n. 2, 2003, p. 227-256.
- [2] K. R. Scherer, “Vocal affect expression: a review and a model for future research”. In: *Psychological Bulletin*, v. 99, n. 2, 1986, p. 143-165.
- [3] T. Johnstone e K. Scherer, “Vocal communication of emotion”. In: M. Lewis e J. Haviland (orgs.) *Handbook of emotions*. 2. ed. Nova Iorque: Guilford, 2000. p. 220-235.
- [4] C. Darwin, *A expressão das emoções no homem e nos animais*. São Paulo: Companhia das Letras, 2009[1872].
- [5] P. Ekman, “An argument for basic emotions”. In: *Cognition & Emotion*, v. 6, n. 3/4, 1992, p. 169 – 200.
- [6] K. R. Scherer, R. Banse, H. G. Wallbott, “Emotion Inferences from Vocal Expression Correlate across Languages and Cultures”. In: *Journal of Cross-Cultural Psychology*, v. 32, n. 1, 2001, p. 76 – 92.
- [7] R. R. Cornelius, “Theoretical approaches to emotion”. In: *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, 2000, Newcastle, Irlanda do Norte, 2000. p. 3-10.
- [8] H. Schlosberg, “Three dimensions of emotion”. In: *Psychological Review*, v. 61, n. 2, 1954. p. 81-88.
- [9] W. da Silva, *Análise fonético-acústica da expressividade de emoções em depoimentos reais*. Dissertação de Mestrado (Linguística) - Universidade Estadual de Campinas, 2014.
- [10] P. A. Barbosa, “Detecting changes in speech expressiveness in participants of a radio program,” *Proceedings of Interspeech 2009 - Speech and Intelligence*. Londres: Causal Productions, 2009. p. 2155-2158.
- [11] P. Boersma, D. Weenink, *Praat: doing phonetics by computer (Versão 5.2.25)*. [Computer program]. Disponível em: <http://www.praat.org>. Acesso em 2011.
- [12] S. Madureira, “The investigation of speech expressivity,” in *Pragmatics and Prosody*, Firenze: Firenze University Press, 2011, p. 101-118.