

A manifestação da emoção na fala: estudo perceptual com falantes nativos e não nativos

(Emotional speech: perceptual study with native and non-native speakers)

Daniel Oliveira Peres¹

¹Faculdade de Filosofia, Letras e Ciências Humanas – Universidade de São Paulo (USP)

danielperes@usp.br

Abstract: This pilot study aims to analyse the expression of emotion in speech by conducting an analysis of production and perception focusing on the debate about the universality of emotional expression in humans, started by the pioneering work of Darwin (1872). The production analysis was based on the evaluation of 8 acoustic parameters: 5 parameters involving intonation and 3 parameters dealing with voice quality. For the analysis, 32 Brazilian Portuguese excerpts were selected and equally divided into anger, fear, joy and sadness. As expected, the degree of agreement between Brazilian subjects was fair to good – $K = 0.73$, $p \sim 0$. In the case of English subjects, the agreement was poor – $K = 0.35$, $p \sim 0$ (FLEISS et al., 2003). The results of one-way ANOVA showed a significant difference between the judgments of emotions. The differences between the means (Tukey HSD) seem to hint at the participants' mistakes.

Keywords: Phonetics; speech perception; emotional speech.

Resumo: Este estudo piloto tem como objetivo analisar a expressão da emoção na fala por meio de análise de produção e percepção, focalizando o debate sobre a universalidade da expressão das emoções nos humanos, iniciado pelo trabalho pioneiro de Darwin (1872). A análise de produção foi baseada em 8 parâmetros acústicos: 5 parâmetros entoacionais e 3 de qualidade vocal. Para a análise, 32 excertos do português brasileiro foram selecionados e divididos igualmente entre raiva, medo, alegria e tristeza. Como esperado, o grau de concordância entre os brasileiros foi de razoável a bom $K = 0.73$, $p \sim 0$, enquanto os ingleses tiveram um desempenho inferior $K = 0.35$, $p \sim 0$ (FLEISS et al., 2003). Os resultados da ANOVA de um fator mostraram uma diferença significativa entre os julgamentos das emoções. A diferença entre as médias (Tukey HSD) parece ser um indício que reflete os erros dos participantes.

Palavras-chave: Fonética; percepção de fala; fala emotiva.

Introdução

Uma das questões centrais que envolvem os estudos de fala expressiva ou emotiva é o papel que a língua e a cultura¹ desempenham na categorização e na percepção das emoções. Da mesma forma, a busca de características universais na manifestação e percepção das emoções é um fator importante para os estudos na área da ciência afetiva.

Darwin (1965[1872]) argumenta que a manifestação das emoções e o seu reconhecimento fazem parte de uma herança biológica, que pode ser explicada por meio de

¹ Sabendo-se que a definição do termo “cultura” pode variar consideravelmente, neste estudo, o termo cultura será entendido, segundo Tylor (1871), como sendo um “*complex whole which includes knowledge, belief, art, morals, law, custom, and any other capabilities and habits acquired by man as a member of society*” (TYLOR, 1871, p. 1).

três princípios: *princípio dos hábitos associados úteis*, *princípio da antítese* e *princípio das ações diretas do sistema nervoso*. Esses princípios são válidos para os animais, dentre eles os humanos, e independentes de regiões, línguas e sociedades.

Seguindo uma abordagem social das emoções, Russel (1991) afirma que a divisão do mundo em categorias é uma característica humana que é estendida à categorização que cada língua faz das emoções. O que está em questão, nesse tipo de abordagem, não é qual emoção está representada em determinada língua, mas como os falantes de uma mesma língua a percebem.

A partir desses dois tipos de abordagem, este estudo objetiva analisar a percepção da fala emotiva por falantes nativos e não nativos, tentando encontrar pistas acústicas no esforço vocal dispensado para a produção da fala e nas características entoacionais que possam explicar os julgamentos dados pelos participantes. Para tal, trechos de fala espontânea emotiva em português brasileiro foram apresentados a falantes nativos do português e do inglês.

A hipótese que norteia este estudo é:

- As emoções podem ser julgadas por falantes não nativos de maneira semelhante ao julgamento feito pelos falantes nativos.

Questionamentos decorrentes da hipótese:

- Caso seja confirmada a hipótese, quais os fatores que podem explicar o seu resultado?
- Caso seja refutada, quais os caminhos para futuros refinamentos no desenho experimental e na análise dos dados?

Este estudo está dividido em: Emoção – delimitação e abordagens; Por que fala espontânea?; Estudos sobre emoção entre línguas distintas; Metodologia; Resultados e Considerações finais.

Emoção – delimitação e abordagens

Embora as emoções sejam quase sempre imediatamente reconhecidas quando externadas pelos falantes, a delimitação delas não é simples. O número de palavras utilizadas pelas línguas para nomear emoções, sentimentos e atitudes é um bom exemplo das diferenças existentes, variando consideravelmente entre as línguas (WALLACE; CARSON, 1973; HOEKSTRA, 1986; BOUCHER, 1979; LUTZ, 1980; HOWELL, 1981 apud RUSSEL, 1991). O quadro abaixo demonstra quão variável é essa delimitação:

Quadro 1. Número de termos relacionados às emoções em diferentes línguas (adaptado de Russel (1991))

Língua	Número de palavras para emoções	Pesquisadores
Inglês	Mais de 2000	Wallace e Carson (1973)
Holandês	1501	Hoekstra (1986)
Chinês (Taiwan)	750	Boucher (1979)
Malaio	230	Boucher (1979)
Ifalukian ²	58	Lutz (1980)
Chewong ³	7	Howell (1981)

Além disso, o próprio conceito de emoção, embora tido como universal no estudo de Brandt e Boucher (1986), pode não ter um termo correlato válido universalmente.

Do ponto de vista psicológico, as emoções podem ser entendidas como estados ou como processos. No primeiro caso, são tipos de estados mentais que interagem entre si e desencadeiam certos comportamentos e expressões. No segundo caso, a emoção é um processo que se inicia no momento da percepção do estímulo (externo ou interno) até a resposta corporal correlata, e.g., aceleração do pulso cardíaco, aumento da pressão arterial, sudorese etc. Como exemplo, Cosmides e Tooby (2000) citam o medo provocado nos hominídeos por estarem sozinhos durante a noite, correndo o risco de ataque de predadores.

As abordagens teóricas da emoção podem ser divididas em evolutivas, sociais e processos emocionais. Neste estudo, como dito acima, será feito um breve esboço das duas primeiras – evolutivas e sociais –, pois estão relacionadas diretamente aos objetivos deste trabalho.

As teorias evolutivas tratam da maneira como as emoções foram herdadas pelos humanos, podendo ter sido selecionadas nos hominídeos. Essa questão remonta ao trabalho pioneiro de Darwin (1965[1872]), que enumera uma série de traços da expressão emotiva nos homens e nos animais. As emoções, segundo Darwin, podem ser explicadas por meio de três princípios, a saber: i) princípio dos hábitos associados úteis – ações despertadas no organismo que tiveram alguma utilidade em estados de espírito distintos no passado e que, ao menor sinal desses estados no presente, há uma tendência em repeti-las por associação e hábito, mesmo que não tenham função aparente; ii) princípio da antítese – quando um estado de espírito contrário é induzido, há tendência involuntária para a execução de movimentos opostos, mesmo que não nunca tenham sido úteis, e.g. a docilidade dos cães; iii) princípio das ações diretas do sistema nervoso – ações reconhecidas como indicadoras de certos estados de espírito são diretamente atribuídas à constituição do sistema nervoso, desde o início independentes do hábito e da vontade. Com esses princípios postos, Darwin explica como as emoções foram herdadas por sucessivas gerações, sem atentar para o fator linguístico e cultural, por serem esses estágios mais tardios do ponto de vista evolutivo.

Para James (1890)⁴, a emoção é a percepção de diferentes mudanças corporais que denotam um estado emotivo. Assim, as emoções são resultado da percepção das reações

2 Língua falada na ilha de Ifaluk. O termo não foi traduzido.

3 Língua falada na península da Malásia.

4 Embora de maneira independente, Carl Lange (1912 [1885]), em sua teoria, também partiu do mesmo pressuposto de James (1890), ou seja, de que o corpo é desencadeador de emoções. Por essa razão, esse tipo de abordagem ficou conhecido como *James-Lange Theory of Emotion*.

fisiológicas e comportamentais desencadeadas no corpo por determinado evento externo ou interno. Assim como Darwin, essa abordagem leva em consideração a função preservadora da espécie, mas propõe uma ordem inversa na manifestação das emoções. Para James o corpo desencadeia as emoções; para Darwin, o corpo é o lugar onde as emoções são manifestadas.

No outro extremo, estão as teorias que explicam as emoções como construções sociais, ou seja, um produto de culturas distintas e apreendidas pelos integrantes de cada sociedade por meio da experiência. Dentre as justificativas, a discrepância entre os nomes dados às emoções em diferentes línguas parece apontar para uma visão social e local das emoções. Além disso, a manifestação das emoções ocorre, em grande número, em contextos sociais, provocadas por outras pessoas em relações sociais distintas. Russel (1991) conclui que:

[...] people of different cultures and speaking different languages categorize the emotions somewhat differently. The boundaries around the domain appear to vary, as do divisions within the domain. Thus, neither the word emotion nor words for even alleged basic emotions, such as anger and sadness are universal. (RUSSEL, 1991, p. 444)

Outra característica desse tipo de abordagem é a assunção de que as emoções e suas expressões decorrem de valores, expectativas e normas sociais. Averill (2013) define algumas regras da manifestação da raiva, dentre elas a proporcionalidade da reação a um estímulo, ou seja, a reação não pode exceder o necessário para corrigir a situação.

No entanto, para tratar da percepção da emoção por falantes nativos e não nativos é necessária uma abordagem que não se fixe somente em um dos extremos, social ou evolutivo, mas que trate das duas faces do problema.

Com vistas a tratar da fala emotiva considerando esses dois extremos, social e evolutivo, Scherer (2006) propõe uma abordagem baseada nos efeitos *push* e *pull*.⁵ O ganho desse tipo de análise é abarcar não somente o que é geral/universal na fala emotiva, mas também tratar das características linguísticas locais responsáveis pela variação que essas manifestações apresentam em diferentes línguas. O efeito *push* é entendido como a influência da ativação psicofisiológica, ligada fortemente à fala emotiva; ao passo que o efeito *pull* é atribuído às normas culturais de expressão, ligadas à pragmática⁶ da língua. Essa distinção implica que o efeito *push*, marcas biológicas, é diretamente exteriorizado na expressão motora, conseqüentemente, na produção da fala; ao passo que o efeito *pull* é baseado em modelos socioculturais que são apreendidos pelos membros de cada sociedade. Com relação às diferenças entre as culturas, é esperado que haja poucas diferenças entre as culturas quando se trata do efeito *push*, e muitas diferenças quando se trata do efeito *pull*.

Para a análise, Scherer (2006) propõe os princípios de covariação e configuração. O princípio da covariação assume uma relação contínua, mas não necessariamente linear, entre o aspecto emocional e a variação acústica. Por exemplo, a frequência fundamental (F0), qualidade de voz e energia das vocalizações assinalam as expressões de medo, raiva, alegria etc. Esse princípio é definido como mais antigo com relação à evolução da linguagem.

5 Optou-se pelo termo em inglês.

6 O termo pragmática em Scherer (2006) não corresponde exatamente ao nível de análise estudado pela linguística, pois é tomado de maneira geral, indicando características gramaticais amplas da língua.

O princípio de configuração determina que o significado específico transmitido por um enunciado é inferido pelo ouvinte a partir da configuração prosódica como um todo e das escolhas linguísticas. Esse princípio parece determinar as características pragmáticas da fala, como ênfase e tipos de sentenças (SCHERER, 2006, p. 13-14). É um princípio mais recente do ponto de vista evolutivo, mais ligado à emergência da linguagem humana.

Assim, pode-se associar as variáveis contínuas com o efeito push, ligado à exteriorização de estados emocionais, enquanto a configuração de categorias pragmáticas da fala são melhor ligadas ao efeito pull, ligado aos contornos entoacionais e demais tipos de padrões encontrados nas línguas naturais.

Por que fala espontânea?

A maioria dos estudos que tratam da fala expressiva faz uso de sentenças com fala teatral ou outros tipos de elicitación (SKINNER, 1935; COSTANZO; MERKEL; COSTANZO, 1969; WALLBOTT; SCHERER; 1986; SCHERER, 2000; SCHERER; BANSE; WALLBOTT, 2001). No português brasileiro, estudos com fala elicitada foram conduzidos por Colamarco e Moraes (2008) e Vassoler e Martins (2013).

A favor da fala teatral ou elicitada, o controle dos estímulos em sentenças idênticas, pronunciadas nas mais variadas emoções, permite ao experimentador um maior controle das variáveis. Scherer (1981) aponta os problemas encontrados em gravações de fala espontânea, sem intervenção direta do experimentador, afirmando que “naturally recorded emotions are by definition singular cases, both in terms of speaker identity, situation context, and verbal content of utterance” (SCHERER, 1981, p. 204). Com essas características da fala espontânea, fica difícil a separação de quais variáveis estão de fato agindo para configurar a fala expressiva, configurando um problema quanto à ortogonalidade do experimento.

Por outro lado, Roberts (2011) demonstra que a fala teatral pode ser fortemente impregnada de estereótipos, o que pode causar diferenças na produção e, conseqüentemente, na percepção dos estímulos. Scherer (1981) também alerta para esse problema, apontando para o uso equivocado de emoções discretas para o julgamento de estímulos com fala teatral. Nesse último caso, poderá haver uma dupla influência nos resultados, partindo dos estímulos estereotipados e da utilização de rótulos estanques para o reconhecimento das emoções.

A fala espontânea sem a influência do experimentador parece ser a melhor escolha para os estímulos, mesmo que a variabilidade, os contextos de produção e as características pessoais mudem caso a caso. Essas influências podem ser contornadas por meio de técnicas de delexicalização e normalização, por exemplo. É preferível intervir nos estímulos espontâneos para diminuir a sua variabilidade a analisar um estímulo com fala teatral que pode, em alguns casos, ser equivocado desde o momento de sua gravação.

Estudos sobre emoção entre línguas distintas

Para este trabalho, dois estudos foram tomados como referência, a saber, Scherer (2000) e Scherer, Banse e Wallbott (2001). Ambos os estudos trabalharam com fala não

espontânea e baseada em cenários. A utilização de cenários para eliciação das sentenças é uma maneira mais confiável, pois evita a produção de fala emotiva a partir de rótulos verbais, e. g. “produza a sentença X como se você estivesse muito triste”. Com os cenários, os falantes vivenciam situações (vinhetas ou pequenos cenários) que ilustram uma situação emotiva. Então, eles são instruídos a imaginar e atuar/falar como se estivessem vivendo essa experiência (WALLBOTT; SCHERER, 1986).

Um primeiro problema é, como dito acima, a utilização de sentenças-padrão produzidas por atores. Um outro problema está na produção de falas sem sentido – *meaningless multilanguage sentences* (WALLBOTT; SCHERER, 1986) – que, embora sejam fruto de uma metodologia que envolve sentenças formadas por sílabas das línguas dos falantes envolvidos no teste, elas não podem ser reconhecidas por nenhum dos participantes. Ademais, por não ser familiar a nenhum dos participantes dos experimentos de percepção – exceto do ponto de vista das sílabas –, a possibilidade de se observar o papel das línguas como desencadeadoras da percepção da emoção pelos falantes nativos está descartada.

Medidas indiretas de esforço vocal

Para tratar das medidas indiretas de esforço vocal, é necessário definir o que se entende por *voz e qualidade vocal*. Normalmente, esses dois termos são tomados como sinônimos, embora seja esse tipo de entendimento fruto de confusão. Kreiman e Sidtis (2011) definem a voz como portadora de uma realidade física e fisiológica que se refere ao sinal acústico, ao passo que a qualidade de voz está relacionada à percepção, ou seja, à impressão que a voz provoca nos interlocutores (KREIMAN; SIDTIS, 2011, p. 5). Esse tipo de divisão no estudo da voz – produção vs percepção – é análoga à feita entre F0 e *pitch*, mas é importante dizer que esta última relação não é biunívoca. Por conta disso, impressões de *pitch* nem sempre decorrem de variações de F0.

Além dessa divisão, o estudo da voz e da qualidade de voz pode ser feito de maneira estrita, considerando somente as ondas sonoras produzidas pela vibração das cordas vocais, ou seja, somente a fonte é considerada, em termos da Acoustic Theory of Speech Production (FANT, 1960). Nessa abordagem, ficam excluídas, por exemplo, as ressonâncias e as turbulências criadas pelo trato vocal. Para uma abordagem mais abrangente, as demais características da produção da voz são consideradas como, por exemplo, os movimentos dos lábios, da mandíbula e da língua. Dessa forma, a voz pode ser entendida como sinônimo de fala. A mesma divisão – estrita ou abrangente – é válida para o estudo da qualidade de voz, permitindo o estudo de uma característica particular da qualidade de voz percebida ou estudo de características gerais.

Neste trabalho, as medidas de esforço vocal estarão, dentro das possibilidades, relacionadas à força com que o som é gerado na fonte. Por isso, as medidas foram feitas em vogais tônicas (exceto inclinação espectral), permitindo um maior controle das influências do trato vocal. Do ponto de vista perceptual, a análise foi feita de maneira abrangente, pois não foram tratados aspectos específicos da qualidade de voz, mas termos mais gerais, como a nomeação das emoções. Assim, este trabalho aborda a voz de maneira estrita e a qualidade de voz de maneira abrangente.

Entoação – Abordagem com ExProsodia

A proposta de análise de F0 realizada pelo aplicativo ExProsodia® (FERREIRA NETTO, 2010) considera a entoação como uma sequência de tons, iguais ou diferentes, produzidos pela voz durante a fala. A análise baseia-se na hipótese de Ferreira Netto (2006, 2008) de que a entoação pode ser entendida como uma série temporal, ou seja, um conjunto de observações sequenciadas e interdependentes, sendo o resultado da observação feita no momento $t+1$ condicionada ao momento t . Em outras palavras, o valor observado em cada momento de medição depende do valor do momento imediatamente anterior, não permitindo inversão dos valores observados. Assim, a ordem de ocorrência de cada momento deve ser mantida como uma informação inerente ao valor obtido.

A análise empreendida na abordagem com ExProsodia está baseada na proposta de Xu e Wang (1997) de que alguns fatos prosódicos são decorrências de restrições *mecânico-fisiológicas* (componentes estruturadoras) e outros configuram as *intenções expressivas* dos falantes (componentes semântico-funcionais). As intenções expressivas seriam as variações intencionais de F0 com fins linguísticos, i.e., alterações na onda sonora que configuram a intenção de imprimir significados diversos à fala. As restrições mecânico-fisiológicas seriam condicionadas pelas características fisiológicas do falante. A subdivisão dessas duas componentes é dada como: *componentes estruturadoras* — finalização (F) e sustentação (S) —, *semântico-funcionais* — foco/ênfase (E) — e acento lexical (A). A figura abaixo ilustra essa divisão:

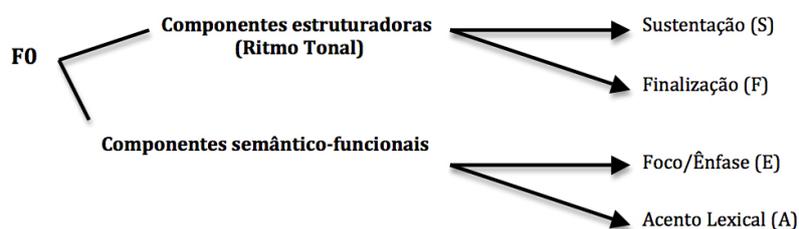


Figura 1. Esquema das componentes de F0

A mensuração de F0 feita pelo ExProsodia é tomada a cada 5 milissegundos, levando em consideração informações predeterminadas pelo pesquisador. As informações requeridas são medidas de F0 (Hz) e intensidade (RMS). O pesquisador também pode colocar limites de duração (ms). A combinação desses parâmetros formará uma *unidade entoacional*, a qual será considerada a porção sonora a ser analisada. Abaixo estão os parâmetros utilizados para as análises feitas neste trabalho:

- **Limiar inferior de frequência fundamental (F0): 50 Hz.**
- **Limiar superior de frequência fundamental (F0): 350 Hz.**
- **Limiar inferior de duração: 20 ms.**
- **Limiar superior de duração: 300 ms.**
- **Limiar de intensidade: 2000 RMS.**

Como resultado, as sentenças são apresentadas, dentre outras possibilidades, da seguinte maneira:

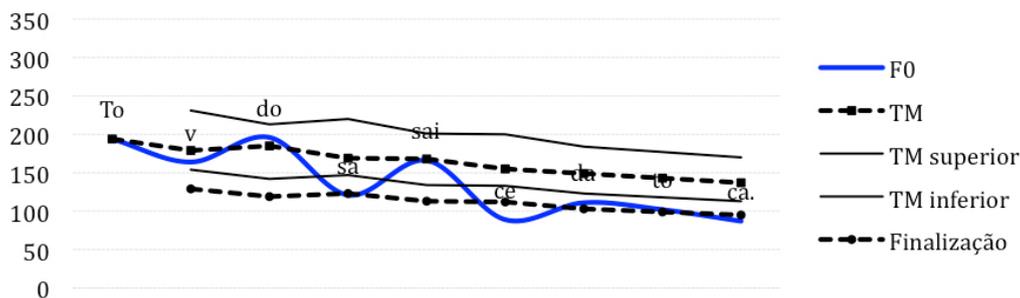


Figura 2. Ilustração de uma sentença analisada pelo ExProsodia

A análise automática tem como resultado parâmetros acústicos que podem ser utilizados na análise entoacional da fala. No caso deste estudo, alguns parâmetros foram selecionados e serão apresentados na próxima seção.

Metodologia

Análise de produção – medidas indiretas de esforço vocal e entoação

Os estímulos utilizados nos experimentos de percepção foram analisados através de medidas indiretas de esforço vocal, a saber: *H1-H2*, *ênfase espectral* e *inclinação espectral*. As medidas de ênfase espectral foram extraídas das vogais tônicas [a] que estavam na região mais próxima possível do centro da sentença. Essa decisão foi tomada para evitar o efeito das bordas da sentença e para se obter um contexto de medida equânime para todas as vogais.

As medidas de ênfase espectral foram extraídas de acordo com os trabalhos de Eriksson, Thunberg e Traunmüller (2001) e Barbosa, Eriksson e Åkesson (2013), *i. e.* $L - L_0$, em que L é a intensidade do espectro da vogal como um todo e L_0 é a intensidade medida até $1,5F_0$. As medidas de *H1-H2* foram extraídas das mesmas vogais utilizadas nas medições de ênfase espectral. As medidas de inclinação espectral foram extraídas das sentenças como um todo.

Para a análise entoacional das sentenças, somente parâmetros de medição de F_0 foram selecionados. Os 32 excertos de fala emotiva foram analisados de maneira idêntica pelo aplicativo. Abaixo estão parâmetros acústicos utilizados:

- Coeficiente de variação de F_0 – *CVF0*
- Tom médio das sentenças – *TM*
- Desvio padrão do tom médio – *DPTM*;
- Assimetria do tom médio – *ATM*
- Coeficiente de variação do tom médio – *CVTM*

Experimento de percepção

Os estímulos utilizados neste estudo foram coletados da internet, no *site* de vídeos www.youtube.com. Foram escolhidos 32 trechos de fala espontânea considerados como fala emotiva. Para a utilização dos estímulos nos experimentos de percepção e na análise

de produção, foi necessária a conversão dos vídeos em arquivos de áudio formato *mp3* (320 kbps).

Com a extração dos 32 arquivos, os estímulos foram apresentados aleatoriamente a falantes do português brasileiros (paulistanos) e do inglês (Inglaterra). É importante salientar que os falantes ingleses não possuem conhecimento de português. Esse cuidado foi tomado para que informações de ordem linguística não pudessem influenciar o resultado da tarefa.

Os estímulos foram intercalados pelo número correspondente de cada sentença, dito nas duas línguas (português e inglês), e separados por um bipe de 0.5s. Foi pedido aos informantes (4 brasileiros e 4 ingleses) para que preenchessem uma planilha, colocando um “X” nas emoções que eles julgavam representar melhor cada um dos estímulos. As emoções utilizadas foram: alegria, tristeza, raiva e medo.

Resultados

Os informantes brasileiros, como esperado, tiveram 90% de acerto, ao passo que os informantes ingleses tiveram um desempenho inferior (66% de acerto). Um teste de inter-rater reliability (Kappa)⁷ foi realizado a fim de avaliar a concordância entre participantes. Como esperado, o grau de concordância entre sujeitos brasileiros foi de razoável a bom – $K = 0,73$, $p \sim 0$. No caso dos participantes ingleses, a concordância foi baixa – $K = 0,35$, $p \sim 0$ (FLEISS; LEVIN; PAIK, 2003).

Um teste de variância (ANOVA – 1 fator) foi feito, envolvendo as emoções (raiva, medo, alegria e tristeza) e os parâmetros entoacionais (CVF0, TM, DPTM, ATM e CVTM). Os resultados do teste para os parâmetros DPTM, ATM e CVTM não foram significativos: $F(3,28) = 0,26$, $p > 0,05$; $F(3,28) = 1,84$, $p > 0,05$ e $F(3,28) = 1,97$, $p > 0,05$, respectivamente. Por outro lado, TM e CVF0 apresentaram resultados significativos: $F(3,28) = 7,97$, $p < 0,05$ e $F(3,28) = 5,55$, $p < 0,05$, respectivamente.

O mesmo procedimento foi adotado para a análise das medidas indiretas de esforço vocal. Somente o parâmetro H1-H2 não foi significativo: $F(3,38) = 1,32$, $p > 0,05$. Os demais parâmetros – inclinação espectral e ênfase espectral – apresentaram resultado significativo: $F(3,38) = 4,92$, $p < 0,05$ e $F(3,38) = 4,25$, $p < 0,05$, respectivamente. Um teste post hoc de Tukey (HSD) foi feito para determinar quais diferenças entre as médias foram encontradas no teste de variância. As diferenças significativas encontradas foram: TM – alegria e medo, tristeza e raiva; CVF0 – tristeza e raiva; ênfase espectral – tristeza e raiva; inclinação espectral – tristeza e raiva.

Considerações finais

A diferença entre as médias parece ser uma característica que reflete os erros dos participantes. Como o experimento perceptual mostrou, os participantes tenderam a reconhecer o medo e a alegria quando se tratava de tristeza e raiva. Esses resultados podem ser entendidos como um julgamento condicionado, principalmente, por emoções não estimulantes (medo e tristeza) e emoções estimulantes (raiva e alegria).

⁷ Toda a análise estatística deste trabalho foi feita por meio do software R.

A alta porcentagem de respostas corretas dadas pelos participantes brasileiros (90%) pode ser explicada pela influência do papel desempenhado pelo léxico e pelo conhecimento pragmático da língua (efeito *pull*). Por outro lado, os participantes ingleses não têm a mesma informação (léxico), mas conseguiram reconhecer as emoções com 66% de precisão. Os parâmetros acústicos analisados neste trabalho podem ser entendidos como pistas que os falantes não nativos utilizam para identificar e rotular os estímulos, sem um conhecimento prévio da língua em questão (efeito *push*).

A utilização de emoções como rótulos estáticos fornecidos aos participantes do experimento pode ter desempenhado um papel importante no índice de acerto dos falantes das duas línguas.

REFERÊNCIAS

AVERILL, J. R. Illusions of anger. In: FOLSON, R. B.; TEDESCHI, J. T. (Ed.). *Aggression and violence: social interactionist perspectives*. Washington, DC: American Psychological Association, 2013. p. 171-192.

BARBOSA, P. A.; ERIKSSON, A.; ÅKESSON, J. Cross-linguistic similarities and differences of lexical stress realisation in Swedish and Brazilian Portuguese. In: ASU, E. L.; LIPPUS, P. (Ed.). *Nordic Prosody*. Proceedings of the XIth conference. Frankfurt am Main: Peter Lang, Tartu, 2013. p. 97-106.

BRANDT, M. E.; BOUCHER, J. D. Concepts of depression in emotion lexicons of eight cultures. *International Journal of Intercultural Relations*, v. 10, p. 321-346, 1986.

COLAMARCO, M.; MORAES, J. A. Emotion expression in speech acts in Brazilian Portuguese: production and perception. In: CONFERENCE ON SPEECH PROSODY, 4, Campinas, 2008. *Proceedings...* Campinas: Unicamp, 2008. p. 717-720.

COSMIDES, L.; TOOBY, J. Evolutionary psychology and the emotions. In: LEWIS, M.; HAVILAND-JONES, J. M. (Ed.). *Handbook of emotions*. 2. ed. New York: Guilford Press, 2000. p. 91-115.

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

DARWIN, C. *The expression of the emotions in man and animals*. University of Chicago Press, Chicago, 1965 [1872].

ERIKSSON, A.; THUNBERG, G. C.; TRAUNMÜLLER, H. Syllable prominence: A matter of vocal effort, phonetic distinctness and top-down processing. *Proceedings of EuroSpeech*, p. 399-402, 2001.

FANT, G. *Acoustic theory of speech production*. The Hague, Netherlands: Mouton, 1960.

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 semântico-funcionais*. Trabalho apresentado no X Congresso nacional de Fonética e Fonologia/ IV Congresso Internacional de Fonética e Fonologia. Universidade Federal Fluminense, Niterói, RJ, 2008.

_____. *Variação de frequência e constituição da prosódia da língua portuguesa*. 2006. Tese (Livre-Docência na Área de Filologia e Língua Portuguesa) – Universidade de São Paulo, São Paulo, 2006.

FLEISS, J.; LEVIN, B.; PAIK, M. *Statistical methods for rates and proportions*. 3. ed. New York: Wiley & Sons, 2003.

JAMES, W. *The principles of Psychology*. New York: Holt, 1890. v. 2.

LANGE, C. G. *The mechanism of emotions*. The Classical Psychologists. Boston: Houghton Mifflin, 1912 [1885] (por Christopher D. Gree). Disponível em: <<http://psychclassics.yorku.ca/index.htm> 1885>.

KREIMAN, J.; SIDTIS, D. *Foundations of voice studies: an interdisciplinary approach to voice production and perception*. Oxford: Wiley-Blackwell, 2011.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em: 2 fev. 2013.

ROBERTS, L. Acoustics effects of authentic and acted distress on fundamental frequency and vowel quality. *Proceedings of The 17th International Congress of Phonetic Sciences (ICPhS XVII)*, 2011, p. 1694-1697.

RUSSEL, J. A. Culture and the categorization of emotions. *Psychological Bulletin*, v. 110, n. 3, p. 426-450, 1991.

SCHERER, K. R. A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology. Implications for Speech. *Proceedings of ICSLP 2000*, Beijing. p. 379-382, 2000.

_____. The affective and pragmatic coding of prosody. In: HUO et al. (Ed.). *ISCSLP-International Symposium of Chinese Spoken Language Processing Proceedings*. Singapore: Springer Berlin Heidelberg, 2006. p. 13-14.

_____. Speech and emotional states. In: DARBY, J. (Ed.) *Speech evaluation in psychiatry*, 1981. p. 189-220.

SCHERER, K. R.; BANSE, R.; WALLBOTT, H. G. Inferences from vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology*, v. 32, p. 76-92, 2001.

SKINNER, E. R. A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness; and a determination of the pitch and force of the subjective concepts of ordinary, soft, and loud tones. *Speech Monographs*, v. 2, p. 81-137, 1935.

TYLOR, E. B. *Primitive culture: researches into the development of mythology, philosophy, religion, languages, art and customs*. London: John Murray Albermale Street, 1871. v. 1.

VASSOLER, A. M. O.; MARTINS, M. V. M. A entoação em falas teatrais: uma análise da raiva e da fala neutra. *Estudos Linguísticos*, v. 42, n. 1, p. 9-18, 2013.

WALLBOTT, H. G.; SCHERER, K. R. Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, v. 51, n. 4, p. 690-699, 1986.

XU, Y.; WANG, Q. E. Component of intonation: what are linguistic, what are mechanical/physiological? *International Conference on Voice Physiology and Biomechanics*, Evanston Illinois, 1997.