



A VELOCIDADE DE FALA COMO PISTA ACÚSTICA DA EMOÇÃO BÁSICA DE RAIVA

Ayane Nazarela Santos de Almeida¹

Miguel Oliveira Jr.²

René Alain Santana de Almeida³

RESUMO

O estudo da expressão acústica da emoção geralmente busca analisar se variáveis prosódicas, como entoação (F0), velocidade de fala, pausas, ritmo, intensidade e duração, são pistas confiáveis para a caracterização dos estados emocionais do falante. O presente trabalho tem como objetivo verificar se existe uma associação no Português do Brasil entre a emoção básica de “raiva” e a variável prosódica “velocidade de fala”, como a literatura frequentemente sugere que há para outros idiomas. O corpus consistiu de fragmentos de fala espontânea gravados a partir de um programa de rádio. Os fragmentos foram selecionados com base em um teste de percepção. Para a análise de produção, apenas trechos que foram identificados por mais de 75% dos participantes do teste perceptual como associados às categorias de “raiva” e “neutro” foram selecionados. Os resultados demonstraram que, para os dados que foram utilizados na análise, há uma redução da velocidade de fala quando os enunciados estão associados com a emoção de “raiva”, se comparado com expressões faladas de forma “neutra” pelo mesmo falante, em oposição ao que a literatura frequentemente indica para outras línguas.

PALAVRAS-CHAVE: velocidade de fala; raiva; fala espontânea.

ABSTRACT

The study of the acoustic expression of emotion is, in general, the analysis of whether prosodic variables such as intonation (F0), speech rate, pauses, rhythm, intensity and duration, are reliable clues for the characterization of the emotional states of the speaker. The present paper aims to verify whether an association exists in Brazilian Portuguese between the basic emotion of “anger” and the prosodic variable “speech rate”, as the literature often suggests that there is for other languages. The corpus consisted of fragments of spontaneous speech recorded from a radio program. The fragments were selected on the basis of a perceptual test. For the production analysis, only excerpts that were identified by more than 75% of the participants of the perceptual test as associated to the categories “anger” and “neutral” were selected. The results demonstrated that, for the data that were used for the analysis, there is a general reduction in speech rate when utterances are associated with the emotion of “anger”, if compared to utterances spoken in a “neutral” mode by the same speaker, contrary to what literature often indicates for other languages.

KEYWORDS: speech rate; anger; spontaneous speech.

1 Universidade Federal de Alagoas, ayanesantos@hotmail.com

2 Universidade Federal de Alagoas, miguel@fale.ufal.br

3 Universidade Federal de Sergipe/Universidade Federal de Alagoas, renealain@hotmail.com



Introdução

As emoções humanas podem ser expressas de várias formas no momento da comunicação: através da expressão facial e corporal do falante, do conteúdo verbal do enunciado e das características acústicas da fala, observadas através do comportamento de parâmetros prosódicos, tais como a intensidade, a duração, a frequência fundamental (F0) e a qualidade de voz.

Todos esses parâmetros, acústicos e visuais, colaboram para a inteligibilidade da fala e promovem a compreensão da expressão de emoções na comunicação humana, conforme podemos verificar em Swerts e Krahmer (2010) e Ekman (2011), por exemplo. Schröder (2003), no entanto, considera que as características prosódicas acústicas sozinhas podem garantir o reconhecimento das emoções expressas pelos seres humanos.

De acordo com Douglas-Cowie et al. (2003), aplicações satisfatórias do mundo real, especialmente as baseadas na interação humano-computador, dependem de um conhecimento aprofundado acerca das formas como as pessoas expressam emoção. A fim de contribuir para a melhoria dos sistemas de reconhecimento e síntese de fala, Scherer (2000) verificou ser de vital importância para a tecnologia de fala determinar as alterações vocais produzidas por fatores emocionais em várias línguas e culturas. Assim, o autor analisou a emoção e reconhecimento de fala em nove países de três continentes diferentes (Alemanha, Suíça, Inglaterra, Holanda, Itália, França, Espanha, Estados Unidos e Indonésia) por meio de um teste de percepção utilizando gravações de sentenças de conteúdo livre em alemão, enunciadas por atores profissionais expressando cinco emoções (medo, nojo, alegria, tristeza e raiva). Seus resultados demonstraram que aspectos segmentais e supra-segmentais contribuem significativamente para a produção e a percepção das categorias emocionais.

A classificação e o conceito de emoção são, no entanto, bastante controversos na literatura (COWIE; CORNELIUS, 2003). Paeschke (2003) afirma que as definições de emoção são tão variadas quanto as pesquisas que a abordam. A classificação das emoções é particularmente controversa por causa de sua característica espontânea, involuntária e imprevisível.

Apesar da falta de consenso geral na literatura, há aspectos básicos comuns que vários trabalhos partilham sobre o assunto, como o fato de que as emoções são quase sempre direcionadas para o objeto, que são ativadas por estímulos internos ou externos e que elas consistem de estados momentâneos de pessoas (PAESCHKE, 2003). Também parece haver um consenso entre os pesquisadores de que a emoção pode ser dividida em três categorias básicas: tristeza, alegria e raiva (PAESCHKE, 2003; SAWAMURA et al., 2007). Entretanto, há autores que além dessas três categorias básicas, acrescentam em suas pesquisas outras emoções tais como: medo (MORAES; RILLIARD, prelo) e além desta, nojo (SCHERER, 2000), por exemplo.

Cada uma das emoções, conforme classificado pela literatura, está associada com características prosódicas específicas que as individualizam. Paeschke (2003), Leon (1993)

e Bezooyen (1984), em estudos das línguas alemã, francesa e holandesa, respectivamente, procuraram associar padrões emocionais nesses idiomas com parâmetros acústicos, tais como frequência fundamental, alcance vocal, intensidade e duração total do enunciado. Em todos esses estudos, os autores observaram que a tristeza é geralmente caracterizada por uma F0 mais baixa, uma voz mais baixa e uma velocidade mais lenta, enquanto que a raiva e a alegria têm uma F0 superior, uma voz mais alta e uma velocidade de fala rápida.

Pereira (2009), por outro lado, ressalta que nem sempre os falantes de uma língua expressam emoções da mesma forma, com os mesmos níveis de ativação. Então, cada uma das emoções básicas podem gerar diferentes níveis de ativação no falante, dependendo da sua manifestação. Por exemplo, Paeschke (2003) observou que a irritação provoca um baixo nível de ativação no falante, enquanto que a fúria o afeta mais intensamente.

Scherer (1986) afirma que pode haver diferenças muito sutis na forma como se manifestam emoções básicas em cada falante. O autor, no entanto, não faz referência a diferentes emoções, mas a subtipos da mesma emoção. Dessa forma, considera “fúria” e “irritação”, por exemplo, como sendo duas manifestações diferentes da mesma emoção: raiva.

De acordo com Douglas-Cowie et al. (2003), muitos estudos linguísticos acerca das emoções e dos estados a elas relacionados não apresentam condições de comparabilidade já que utilizam distintos procedimentos metodológicos, há diferentes classificações das categorias de emoção, e também variações no que diz respeito ao tipo de estímulo utilizado no estudo: real ou atuado. Outros, porém, parecem sugerir diferenças significativas na expressão vocal de emoção: de falante para falante, de cultura para cultura, para diferentes situações e sexos. Os autores salientam que as comparações entre línguas e culturas são limitadas, mas sugerem diferenças substanciais.

Ainda há poucos estudos sobre as características prosódicas de emoções e eventos relacionados com emoção para o Português Brasileiro (PB). Pereira (2009), por exemplo, investigou como a expressão de três emoções primárias afeta os contornos melódicos característicos de quatro atos de fala no PB falado no Rio de Janeiro. Para testar isso, a autora utilizou uma interação entre quatro atos de fala (afirmação, pergunta, solicitação e ordem) e quatro emoções (três emoções principais: tristeza, alegria e raiva, além da forma “neutra”). Ela procurou responder se essas duas categorias - atos de fala e emoção - são realmente dimensões prosódicas independentes ou se a interação entre elas provoca alterações substanciais nos padrões entoacionais que têm sido encontrados em estudos que abordam apenas atos de fala.

A autora encontrou em seus resultados que atos de fala e padrões emocionais parecem ser categorias independentes em termos de produção. No que diz respeito à percepção, no entanto, existe alguma sobreposição entre as duas categorias, já que os atos de “afirmação” e “pergunta” pareceram aceitar mais facilmente qualquer padrão emotivo enquanto que nos atos de “solicitação” e “ordem” os participantes apresentaram uma dificuldade maior no

reconhecimento da emoção expressa. Ela concluiu que entoação e qualidade de voz devem ser vistas como categorias complementares, ambas necessárias para o reconhecimento de estados emocionais do falante. Entretanto, vale salientar que a autora analisou enunciados produzidos por dois atores da região sudeste do Brasil. A maioria dos poucos estudos que abordam a prosódia de emoções no PB baseia-se nos dialetos do sul e sudeste, e utilizam enunciados monitorados, produzidos por atores.

Barbosa (2009) apresentou um método, baseando-se em análises estatísticas, para mensurar mudanças na expressividade da fala do participante mediante a eventos criados pelo seu interlocutor (o radialista) durante conversa telefônica. O autor utilizou como corpus para seu estudo gravações de um programa de rádio chamado Programa do Chupim, transmitido pela Rádio Metropolitana de São Paulo, na região sudeste do Brasil. Ele concluiu que, analisados cronologicamente, os enunciados revelam claras mudanças na expressividade, e que tais mudanças podem ser identificadas de forma confiável a partir de parâmetros acústicos. Entretanto, o autor salienta que seus resultados estão relacionados para os casos que chamou de propensão à ação, já que a ação do apresentador provoca uma série de reações imediatas no participante, por isso sugere estudos que englobem outras dimensões, incluindo empatia, por exemplo, que está mais relacionada com simpatia e envolvimento.

Santos (2013) analisou a prosódia e a expressividade da fala através de emoções expressas em um poema falado por um ator profissional do sudeste do Brasil. Vassoler e Martins (2013) investigaram o contorno melódico associado à emoção da raiva em discursos teatrais, com base em amostras de três atrizes profissionais do sudeste do Brasil, coletadas em condições de laboratório. Reis e Martins (2011) também analisaram fala monitorada com o objetivo de melhorar a fala sintetizada produzida por um sistema TTS (texto para fala), acrescentando-lhe a informação emocional a partir de uma perspectiva acústica. O modelo usado por eles foi baseado em um dialeto do sudeste do Português do Brasil.

Não está claro, no entanto, até que ponto a fala monitorada reflete verdadeiramente uma emoção que é caracterizada como involuntária e imprevisível. Esta é, na verdade, conforme Swerts e Hirschberg (2010), uma das limitações mais importantes em estudos que envolvem fala e emoção. Nesse sentido, Scherer (2003) enfatizou que altos níveis de ativação não são frequentemente encontrados no caso das emoções desencadeadas a partir de condições de controle.

Velocidade de fala

Uma das primeiras referências de investigação empírica sobre velocidade de fala remonta a Cattell (1886), que resume as suas conclusões do seguinte modo:

[...] leva cerca de duas vezes mais tempo para ler (em voz alta, tão rápido quanto possível) palavras que não têm qualquer conexão como letras que

formam palavras [...]. Quando uma passagem é lida em voz alta a uma velocidade normal, quase o mesmo tempo é gasto para cada palavra como quando palavras que não têm nenhuma conexão são lidas o mais rápido possível.

O estudo em questão abordou vários idiomas, como o inglês, o francês, o alemão, o italiano, o latim e o grego, uma vez que o autor também estava preocupado em estudar a velocidade de fala empregada pelo mesmo falante enquanto falava uma língua estrangeira. De acordo com Cattell (1886), a velocidade na qual uma pessoa fala uma língua estrangeira é determinada pela familiaridade que a pessoa tem com a língua: quanto mais familiar a língua for para a pessoa, mais rápida será a fala.

Beer (1910) foi talvez o primeiro a introduzir uma metodologia objetiva para o estudo da velocidade na fala. Usando um método bastante peculiar de medida que dependia de marcação com fuligem a partir de chamas, ele comparou palavras polissilábicas com palavras monossilábicas, usando a sílaba por segundo como unidade de medida - uma unidade amplamente empregada hoje. Em seu experimento, ele demonstrou que, para uma dada passagem contendo a mesma quantidade de sílabas, palavras polissilábicas são lidas mais rapidamente do que as palavras monossilábicas. Ele alega que isso se deve à quantidade de significado que é realizado em passagens contendo sílabas monossilábicas: quanto maior a quantidade de significado que é transmitida em um enunciado, mais tempo o falante vai demorar para pronunciá-lo.

O conteúdo informacional também é considerado um fator determinante para o estabelecimento da velocidade de fala. Fröschels (1920), por exemplo, comparou a repetição de sílabas sem sentido com a produção de sílabas articuladas dentro de palavras e descobriu que sílabas sem sentido são muitas vezes produzidas com uma velocidade mais lenta do que aquelas que fazem parte de palavras reais. Ele também notou que as pessoas tendem a ser afetadas pela curiosidade quando lêem algum conteúdo desconhecido, o que resultaria em uma aceleração deliberada da fala, como resultado desta curiosidade.

Olson e Koetzle (1936) introduziram considerações importantes para o estudo da velocidade de fala, ao investigar a fala das crianças do jardim de infância. Eles concluíram que, enquanto os meninos tendem a falar menos do que as meninas, eles fazem isso em uma velocidade mais rápida. Fairbanks e Hoaglin (1940) consideraram outros fatores externos. Segundo esses autores, as emoções como raiva, medo e indiferença estão intimamente relacionadas com a velocidade de fala rápida, enquanto que desprezo e sofrimento estão associadas a velocidade lenta.

Essen (1949) argumenta que a velocidade de fala reflete processos psicológicos: “A variação de velocidade é, obviamente, controlada em primeira instância pelo significado das unidades individuais do enunciado; quanto mais significado está presente em uma palavra, maior será o aumento da tensão psicológica e física, e mais lenta a velocidade de fala” (ESSEN,

1949, p. 325).

Com base no panorama apresentado acima, pode-se concluir que há uma tradição histórica nos estudos sobre velocidade de fala de relacionar a sua variação com aspectos semânticos e emocionais, associados às diferentes atividades de fala. O estudo sobre velocidade de fala, como um correlato acústico da emoção, ainda é incipiente para o PB, como revisão de literatura apresentada acima sugere. A maioria dos poucos estudos realizados até agora se concentram em características entoacionais na expressividade da fala.

Um dos poucos estudos para PB que aborda o parâmetro acústico de velocidade de fala como um correlato acústico de emoção é o de Pereira (2009). De acordo com suas descobertas, não há uma relação direta entre a duração e diferentes emoções básicas: a duração pode variar como uma particularidade de cada falante.

No presente estudo, objetivamos investigar se a velocidade de fala é, no PB, um correlato acústico confiável de uma emoção humana básica: a raiva. A justificativa para este estudo reside na escassez de tal investigação, apesar do fato de que a literatura muitas vezes relaciona o parâmetro acústico de velocidade de fala com emoções humanas (PAESCHKE, 2003; LEON, 1993; BEZOOYEN, 1984; PEREIRA, 2009).

As relações entre prosódia e emoção no PB foram muito pouco exploradas até este ponto. Os poucos trabalhos que abordam a questão o fazem quase que exclusivamente com emoção atuada e baseiam-se em dados representativos dos dialetos falados na região sudeste do Brasil. A fim de contribuir com a pesquisa ainda incipiente na prosódia emocional no PB, a presente investigação utilizou material espontâneo, representante de um dialeto que tem sido negligenciado até agora neste tipo de investigação: aquele falado na região nordeste do Brasil.

Métodos

Os dados utilizados neste estudo consistem em dezoito pequenos fragmentos de gravações de fala espontânea, seis fragmentos de cada falante com uma duração média de 5,5 segundos cada. Tais fragmentos foram extraídos de um programa de rádio chamado “A Hora do Mução”. É um programa popular de pegadinhas⁴, em que o radialista Rodrigo Vieira Emerenciano encarna o personagem Mução. O radialista telefona para pessoas de qualquer estado da região nordeste do Brasil, a partir de sugestões feitas por amigos e familiares da vítima, a fim de irritá-los, fazendo referência a uma característica física ou um apelido da vítima. Para Barbosa (2009), uma das vantagens desse tipo de interação, no programa de rádio, é a possibilidade de obtenção de elevados níveis de ativação de reações emocionais, devido aos eventos críticos introduzidos pelo apresentador do programa.

4 “Encenação filmada secretamente onde os participantes são surpreendidos com situações cômicas, constrangedoras, provocatórias ou insólitas (Dicionário Priberam da Língua Portuguesa, 2008-2013, <http://www.priberam.pt/dlpo/pegadinha>, consultado em 23-09-2015).

Nas gravações, existe uma clara diferença entre a fala dos participantes na primeira metade dos telefonemas e na segunda metade, quando o motivo das ligações é conhecido (ou seja, a pegadinha está em curso). Na segunda metade dos telefonemas, a fala de quem atendeu ao telefone está claramente relacionada à expressão de raiva.

As gravações desse programa de rádio são, portanto, um material excelente para o tipo de análise proposto aqui, porque é possível fazer comparações, para cada falante, de recursos prosódicos associados à fala neutra com aqueles associados à expressão de raiva.

Além disso, ao contrário do que acontece com a maior parte dos dados utilizados em pesquisa sobre a relação entre prosódia e emoção, as gravações utilizadas na presente pesquisa refletem a fala associada com a emoção ocorrida espontaneamente, o que torna os dados ecologicamente válidos. Também deve ser salientado que a fala ao telefone é ideal para o estudo de pistas acústicas da emoção, porque, ao contrário das interações face a face, ela não apresenta qualquer informação visual que poderia afetar a análise (DOUGLAS-COWIE et al., 2003; PAULMANN; PELL, 2011), já que gestos e outras informações visuais (movimentos de cabeça e sombrancelhas, por exemplo) influenciam na percepção da emoção expressa (SWERTS e KRAHMER, 2010).

Os trechos das gravações foram selecionados a partir de um teste de percepção realizado com quarenta estudantes de graduação em Letras da Universidade Federal de Alagoas. Os dados que foram utilizados neste teste de percepção consistiram em fragmentos de gravações disponível no CD “Pegadinhas do Mução” (EMERENCIANO, 2010), que foram selecionados com base nos seguintes critérios: (i) qualidade do som (inteligível e sem ruído), (ii) sexo do falante (todos os excertos foram extraídos a partir de gravações que apresentam apenas homens, de modo a evitar a inclusão de uma variável que poderia confundir a interpretação dos resultados), e (iii) o conteúdo da informação lexical (os trechos foram selecionados levando em consideração seu conteúdo: por razões óbvias, não deveriam oferecer pistas lexicais evidentes de raiva - ou qualquer outra emoção a ela relacionada). Em relação a este último critério, no entanto, deve-se salientar que estudos anteriores demonstraram que o processamento da prosódia e o processamento semântico linguístico não ocorrem simultaneamente, provavelmente porque ambos dependem de mecanismos neurais parcialmente dissociados (FRIEDERICI; ALTER, 2004; ROSS; MONNOT, 2008).

As transcrições dos trechos selecionados a partir de gravações de três pegadinhas, feitas com três homens diferentes, todos falantes do dialeto de Recife (Pernambuco), juntamente com seus arquivos de áudio correspondentes foram apresentadas em slide-show, de forma aleatória, para os participantes do teste perceptual. Para cada trecho, o participante tinha de identificar uma emoção relacionada ao trecho que escutou. As opções foram as três emoções básicas, também utilizadas por Pereira (2009): “raiva”, “alegria” e “tristeza”, bem como a forma “neutra”. Os

participantes foram instruídos a ouvir cada estímulo quantas vezes eles desejassem antes de responder o formulário.

Para a análise de produção, apenas trechos que foram identificados por mais de 75% dos participantes do teste perceptual como associados às categorias de “raiva” e “neutro” foram selecionados. Como os participantes tinham quatro categorias para escolher uma, o nível de concordância entre eles devido ao acaso seria de 25%.

Há uma variedade de unidades de medida que são empregadas na investigação sobre velocidade de fala: sons por unidade de tempo (ESSEN, 1949; HEINITZ, 1921; HEGEDUS, 1957; FONAGY; MAGDICS, 1960), palavras por unidade de tempo (BRUBAKER, 1972; BARIK, 1977; GROSJEAN; DESCHAMPS, 1975; GROSJEAN; DESCHAMPS, 1972; SCOLLON, 1981), sílabas por unidade de tempo (GROSJEAN; DESCHAMPS, 1972; ABERCROMBIE, 1967; MEINHOLD, 1972) e, finalmente, batidas por unidade de tempo (SCOLLON, 1981). Esta multiplicidade de unidades de medidas empregada na literatura, não só reflete uma falha metodológica séria - como discutido em Meinhold (1972), mas também torna a tarefa essencial da comparação dos resultados entre os vários estudos impossível.

A unidade de medida mais adequada para velocidade de fala, de acordo com muitos autores, é a de sílabas por unidade de tempo (UHMANN, 1992; O’CONNELL; KOWAL, 1972). Abercrombie (1967), por exemplo, define velocidade de fala como “velocidade de sucessão sílabica”. Esta é a unidade adotada por Grosjean e Deschamps (1972), Uhmman (1992), Blaauw (1995), Fon (1999), Grosz e Hirschberg (1992), Goldman-Eisler (1961), Van Donzel (1999) e Wood (1975), para citar alguns. Cumpre notar, todavia, que esta unidade de medida não está livre de problemas, sendo talvez o principal deles o fato de não levar em consideração os processos que podem resultar em omissões de sílabas, fenômeno bastante comum na fala rápida, como a assimilação e o apagamento silábico (O’CONNELL E KOWAL, 1972).

O presente estudo optou por essa medida, que é mais utilizada em pesquisa temporal de fala para fins de comparabilidade. Reconhecemos os problemas relacionados a esta escolha, mas assumimos que eles não são tão graves a ponto de invalidar a interpretação dos dados. A velocidade de fala será analisada neste estudo, portanto, utilizando a medida de sílabas fonéticas por segundo.

Resultados

O Gráfico 1 abaixo mostra as médias de velocidade de fala dos excertos que foram selecionadas para análise, com base no teste de percepção. Está discriminado por emoção (“neutro” e “raiva”) e falante (um total de três).

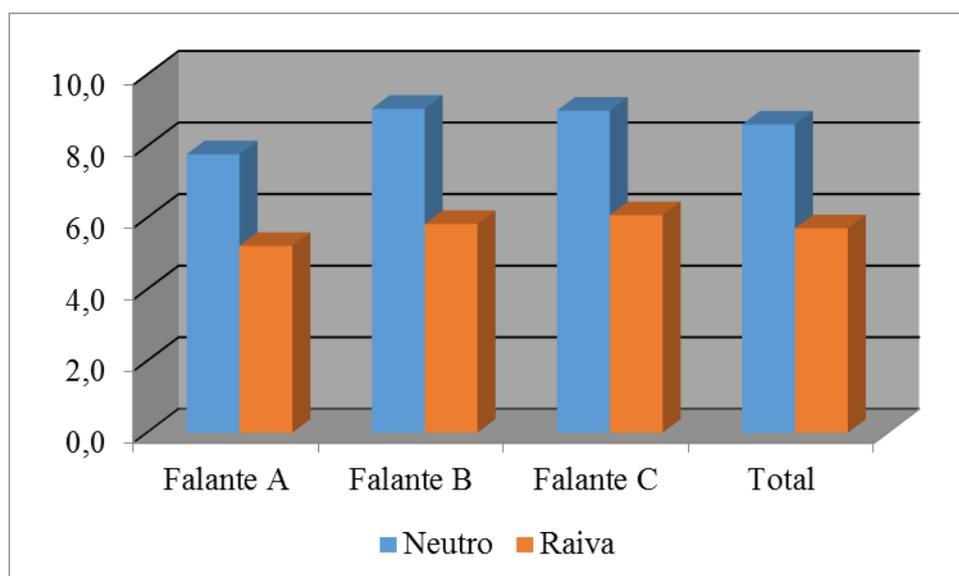


Gráfico 1: Médias de velocidade de fala dos enunciados “neutro” e “raiva”.

Nós usamos a média para comparação entre enunciados associados a cada emoção porque a diferença entre ela e a mediana foi insignificante, o que nos permite afirmar que não há nenhuma distorção significativa nas médias das velocidades encontradas nos trechos analisados e, portanto, pode ser considerada uma medida confiável para esses dados.

Os resultados mostrados no Gráfico 1 indicam claramente que os falantes empregaram uma velocidade de fala mais lenta, em expressões associadas à emoção de raiva, se comparado com aquelas consideradas neutras, tanto individualmente quanto quando considerado o total da amostra. A velocidade de fala dos trechos identificados como “neutro” foi, em média, três sílabas por segundo (ou, cerca de 50%) mais rápida do que nos que foram identificados como associados à emoção de raiva.

Uma série de amostras pareadas e o teste t para um p-valor bicaudal foram conduzidos para avaliar a significância das diferenças na velocidade de fala entre enunciados “neutros” e enunciados associados à expressão de “raiva” para cada falante e para a amostra total. A Tabela 1 abaixo mostra os resultados desses testes:

Amostra	Resultados
Falante A	[t(5) = 7.02, p < 0.0009]
Falante B	[t(5) = 11.08, p < 0.0001]
Falante C	[t(5) = 6.77, p < 0.0011]
Total da amostra	[t(17) = 13.86, p < 1.08E-10]

Tabela 1. Resultados de teste-t, discriminados por falante e pelo total da amostra.

A Tabela 1 evidencia que para todos os falantes, as velocidades de fala diferem significativamente entre as duas condições: “neutro” e “raiva”. Ela também indica que as diferenças são significativas quando consideramos a amostra como um todo. Além disso, os

resultados da Tabela 1 mostram uma baixa probabilidade de os resultados serem decorrentes do acaso, como é indicado pelos valores de p.

Discussão

De acordo com a literatura, a emoção reconhecida como “raiva” é acusticamente caracterizada por velocidade de fala rápida, alta intensidade, alto nível de pitch, muita variabilidade nos valores de F0, subida do contorno melódico, onsets rápidos, e irregularidade microestrutural (PAESCHKE, 2003; LEON, 1993; BEZOOYEN, 1984; PEREIRA, 2009).

O que os resultados gerais da análise de velocidade de fala aqui relatados sugerem é que existem possíveis dissonâncias entre o que a literatura estabelece como um padrão prosódico associado com a emoção de “raiva”, e o que foi identificado como um padrão para a mesma emoção no PB, com base no material de fala espontânea. Os resultados demonstram que há uma redução geral na velocidade de fala quando os enunciados estão associados com a emoção básica de “raiva”, se comparado com expressões faladas em um modo “neutro” pelo mesmo falante. A análise estatística indica que esta diferença não é decorrente do acaso ou devido às características individuais dos falantes, porque as amostras de cada falante foram analisadas individualmente, bem como em grupo, e características muito semelhantes no que diz respeito a esta variável de velocidade de fala na expressão de raiva foram encontradas.

Sabe-se que as emoções manifestam-se de forma contínua, em dimensões diferentes, dependendo do nível de ativação/estimulação (PEREIRA, 2009). A emoção de “raiva”, por exemplo, pode ser expressa numa escala que varia de uma leve “irritação” à “ira”.

Neste trabalho, no entanto, não consideramos os variados níveis de ativação, mas sim a emoção básica “raiva”. É claro que estamos cientes de que, dependendo de onde no *continuum* uma determinada emoção está localizada, sua característica acústica correspondente pode variar (SCHERER, 1989). A despeito disso, a velocidade de fala mais rápida é comumente associada a qualquer nível de ativação na escala de “raiva” (SCHERER, 1986).

A diferença em termos de velocidade de fala como um correlato acústico da emoção de “raiva” aqui relatada, em comparação com os resultados apresentados em outros estudos, pode ser interpretada como um resultado das especificidades dos dados que foram utilizados para a análise: material espontâneo, derivado de fala ao telefone, proferida por falantes masculinos de dialetos pouco estudados do PB.

Conforme Douglas-Cowie et al. (2003) salientam, variações nesse tipo de estudo podem refletir diferenças na expressão vocal de emoção, de falante para falante, de cultura para cultura, e em todas as situações. A fim de descobrir se este é sistematicamente o caso, uma investigação mais aprofundada precisa ser feita, com um corpus maior, contemplando todas as variáveis possíveis que são potencialmente importantes para qualquer tentativa de generalização.

Conclusão

De acordo com Ekman (1992), a comunicação de emoções é crucial para as relações sociais e de sobrevivência. É, portanto, essencial compreender todos os aspectos relacionados a ela, incluindo suas propriedades acústicas. Tem sido sugerido, no entanto, que os sinais de emoção na fala não são consistentes em todos os indivíduos e ocasiões, o que necessita de descrições sistemáticas de suas propriedades para comparações adequadas entre línguas e culturas.

A análise aqui apresentada nos permitiu inferir que, em dados de fala espontânea representativos da variedade do português falado no nordeste do Brasil, não há uma associação entre a emoção básica de “raiva” e um aumento na velocidade de fala, como a literatura indica frequentemente para outras línguas e variedades. Em vez disso, muitos trechos, identificados como sendo representativos da emoção “raiva” pelos participantes do teste de percepção, foram enunciados com uma velocidade de fala mais lenta do que nos trechos percebidos como “neutro”.

Ao contrário da maioria dos estudos sobre as relações entre prosódia e emoção, o presente estudo optou por utilizar o material de fala espontânea para a análise, porque, como Scherer (1986), acreditamos que se a pesquisa nessa área continuar a ser quase exclusivamente baseada em dados artificiais, que simulam emoções, não será possível compreender em que medida estamos operando dentro de um sistema fechado de regras de associação que podem estar apenas indiretamente relacionadas à biologia subjacente da emoção.

O presente trabalho apresenta-se como uma contribuição para a caracterização acústica de padrões emocionais no PB. Planejamos, para investigação futura, ampliar o corpus, de modo a refletir outras variáveis, e incluir outros parâmetros prosódicos na análise, tais como a variação da frequência fundamental, a pausa e a intensidade. Testes de percepção com amostras estilizadas do corpus também estão planejadas, pois entendemos que é fundamental descrever quais parâmetros prosódicos são relevantes do ponto de vista da percepção.

Referências

ABERCROMBIE, D. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press, 1967.

BARBOSA, P. A. “Detecting changes in speech expressiveness in participants of a radio program”. *Proceedings of Interspeech*, pp. 2155-2158, 2009.

BAKIK, H. C. “Cross-linguistic study of temporal characteristics of different speech materials”. *Language and Speech* 20, pp. 116-126, 1977.

BEER, M. “Die Abhängigkeit der Lesezeit von psychologischen un sprachlichen Faktoren”. *Zeitschrift für Psychologie* 56, pp. 264-298, 1910.

- BEZOOYEN, R. *Characteristics and recognizability of vocal expression of emotion*. Dordrecht: Foris, 1984.
- BLAAUW, E. *On the perceptual classification of spontaneous and read speech*. Research Institute for Language and Speech, Utrecht University, 1995.
- BRUBAKER, R. S. "Rate and pause characteristics of oral reading". *Journal of Psycholinguistic Research* 1(2), pp. 141-147, 1972.
- CATTELL, J. M. "The time it takes to see and name objects". *Mind* 11, pp. 63-65, 1886.
- COWIE, R.; CORNELIUS, R. R. "Describing the emotional states that are expressed in speech". *Speech Communication* 40(1), pp. 5-32, 2003.
- DOUGLAS-COWIE, E. et al. "Emotional speech: Towards a new generation of databases". *Speech Communication* 40(1), pp. 33-60, 2003.
- EKMAN, P. "An argument for basic emotions". *Cognition & Emotion* 6, pp. 169-200, 1992.
- _____. *A linguagem das emoções: revolucione sua comunicação e seus relacionamentos reconhecendo todas as expressões das pessoas ao redor*. Trad. Carlos Szlak. São Paulo: Lua de Papel, 2011.
- EMERENCIANO, R. V. *Pegadinhas do Muçãõ*. Rio de Janeiro: Polysom. 1 disco compacto: digital, stereo, 2010.
- ESSEN, O. V. "Sprechtempo als Ausdruck psychischen Geschehens". *Zeitschrift für Phonetik* 3, pp. 317-341, 1949.
- FAIRBANKS, G.; HOAGLIN, L. W. "An experimental study of the durational characteristics of the voice during the expression of emotion". *Speech Monographs* 7, pp. 85-90, 1940.
- FON, J. "Speech rate as a reflection of variance and invariance in conceptual planning in storytelling". *Proceeding of the ICPHS*, San Francisco, pp. 663-666, 1999.
- FÓNAGY, I.; MAGDICS, K. "Speed of utterance in phrases of different lengths". *Language and Speech* 4, pp. 179-192, 1960.
- FRIEDERICI, A. D.; ALTER, K. "Lateralization of auditory language functions: A dynamic dual pathway model". *Brain & Language* 89, pp. 267-276, 2004.
- FRÖSCHELS, E. "Untersuchungen über das Sprechtempo". *Monatsschrift für Ohrenheilkunde und Laryngo-Rhinologie* 54, pp. 867-871, 1920.
- GOLDMAN-EISLER, F. "The rate of changes in the rate of articulation". *Language and Speech* 4, pp. 171-174, 1961.
- GROSJEAN, F.; DESCHAMPS, A. "Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation". *Phonetica* 14, pp. 141-148, 1975.
- Diadorim*, Rio de Janeiro, vol.17, número 2, p. 198-211, 2015.

_____. “Analyse des variables temporelles du français spontané”. *Phonetica* 26, pp. 126-156, 1972.

GROSZ, B.; HIRSCHBERG, J. “Some intonational characteristics of discourse structure”. *Proceeding of the International Conference on Spoken Language Processing*, Banff, pp. 429-432, 1992.

HEGEDÜS, L. “Sprechtempoanalysen im Ungarischen”. *Zeitschrift für Phonetik* 10, pp. 8-20, 1957.

HEINITZ, W. “Die Bewertung der Dauer in phonetischen Aufnahmen”. *Vox* 153, 1921.

LEON, P. *Précis de phonostylistique parole et expressivité*. Paris: Nathan, 1993.

MEINHOLD, G. “Allgemeine Probleme der Sprechgeschwindigkeit”. *Zeitschrift für Phonetik* 25, pp. 492-505, 1972.

MORAES, J.; RILLIARD, A. “Prosody and Emotion in Brazilian Portuguese”. In.: ARMSTRONG, M.; HENRIKSEN, N.; VANRELL, M. (org.). *Interdisciplinary approaches to intonational grammar in Ibero-Romance intonation, Issues in Hispanic and Lusophone Linguistics*, no prelo.

O’CONNELL, D. C.; KOWAL, T. D. “Cross-linguistic pause and rate phenomena in adults and adolescents”. *Journal of Psycholinguistic Research* 1, pp. 155-164, 1972.

OLSON, W. C.; KOETZLE, V. S. “Amount and rate of talking of young children”. *Journal of Experimental Education* 5, pp. 175-179, 1936.

PAESCHKE, A. *Prosodische Analyse emotionaler Sprechweise*. Berlin: Logos Verlag, 2003.

PAULMANN, S.; PELL, M. D. “Is there an advantage for recognizing multi-modal emotional stimuli?”. *Motivation and Emotion* 35, pp. 192-201, 2011.

PEREIRA, M. C. C. *A expressão das emoções em atos de fala no português do Brasil: produção e percepção*. 2009. Dissertação (Mestrado em Letras Vernáculas) – Faculdade de Letras, UFRJ, Rio de Janeiro, 2009.

REIS, B. F.; MARTINS, V. V. “Síntese prosódica da fala em português do Brasil”. *X SBAL*, pp. 1185-1188, 2011.

ROSS, E. D.; MONNOT, M. “Neurology of affective prosody and its functional–anatomic organization in right hemisphere”. *Brain & Language* 104, pp. 51-74, 2008.

SANTOS, I. *Expressividade da Fala: o desvelar da locução de um poema a partir da Análise Acústica e da filosofia de Spinoza*. 2010. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem), Pontifícia Universidade Católica, São Paulo, 2010.

SAWAMURA, K. et al. “Common factors in emotion perception among different cultures”. *Proceedings of 16th ICPhS*, Saarbrücken, pp. 2113-2116, 2007.

SCHERER, K. R. “A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology”. *Proceedings of 6th ICSLP*, Beijing, pp. 379-382, 2000.

_____. “Vocal affect expression: a review and a model for future research”. *Psychological Bulletin* 99, pp. 143-165, 1986.

_____. “Vocal measurement of emotion”. In.: PLUTCHIK, R.; KELLERMAN, H. (org.). *Emotion: theory, research, and experience. The measurement of emotion*. New York: Academic Press, 1989, pp. 233-260.

_____. “Vocal communication of emotion: A review of research paradigms”. *Speech Communication* 40, pp. 227-256, 2003.

SCHRÖDER, M. “Experimental study of affect bursts”. *Speech communication special issue speech and emotion* 40, pp. 99-116, 2003.

SCOLLON, R. *Tempo, Density, and Silence: Rhythms in Ordinary Talk*. Center for Cross-Cultural Studies. Fairbanks, University of Alaska, 1981.

SWERTS, M.; HIRSCHBERG, J. “Prosodic predictors of upcoming positive or negative content in spoken messages”. *J. Acoust. Soc. Am.* 128(3), pp. 1337-1345, 2010.

SWERTS, M. G. J.; KRAHMER, E. J. “Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions”. *Journal of Phonetics*, 38(2), pp. 197-206, 2010.

UHMANN, S. “Contextualizing relevance: on some forms and functions of speech rate changes in everyday conversation”. In.: AUER, P.; LUZIO, A. D. *The Contextualization of Language*. Amsterdam: Benjamins, 1992, pp. 297-336.

VAN DONZEL, M. *Prosodic Aspects of Information Structure in Discourse*. 1999. Thesis (PhD) – Faculteit der Geesteswetenschappen, University of Amsterdam, Amsterdam, 1999.

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

WOOD, S. “Speech tempo”. *Working Papers of the Phonetics Laboratory, Lund University* 9, pp. 99-147, 1975.