



**PROGRAMA ASSOCIADO DE PÓS-GRADUAÇÃO EM
FONOAUDIOLOGIA UFPB/UFRN/UNICISAL**



Héryka Maria Oliveira Lima

**BANCO DE VOZES BRASILEIRO NAS VARIAÇÕES DAS EMOÇÕES
(EMOVOX-BR): ELABORAÇÃO E VALIDAÇÃO**

JOÃO PESSOA
2021



UNIVERSIDADE FEDERAL DA PARAÍBA
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
PROGRAMA ASSOCIADO DE PÓS-GRADUAÇÃO EM FONOAUDIOLOGIA

Héryka Maria Oliveira Lima

BANCO DE VOZES BRASILEIRO NAS VARIAÇÕES DAS EMOÇÕES (EMOVOX-BR): ELABORAÇÃO E VALIDAÇÃO

Dissertação apresentada ao Programa Associado de Pós-Graduação em Fonoaudiologia da Universidade Federal da Paraíba – UFPB, Universidade Federal do Rio Grande do Norte – UFRN, e Universidade Estadual de Ciências da Saúde de Alagoas – UNCISAL, como requisito regulamentar para obtenção do título de mestra em Fonoaudiologia.

Linha de Pesquisa: Voz e funções orofaciais
Orientadora: Profa. Dra. Anna Alice Almeida

JOÃO PESSOA
2021

Catálogo na publicação
Seção de Catalogação e Classificação

L732b Lima, Héryka Maria Oliveira.

Banco de vozes brasileiro nas variações das emoções (EMOVOX-BR) : elaboração e validação / Héryka Maria Oliveira Lima. - João Pessoa, 2021.
47 f.

Orientação: Anna Alice F. de Almeida Queiroz.
Dissertação (Mestrado) - UFPB/CCS.

1. Fonoaudiologia. 2. Voz. 3. Emoções. 4. Comportamento. I. Queiroz, Anna Alice F. de Almeida.
II. Título.

UFPB/BC

CDU 616.89-008.434.5(043)



**UNIVERSIDADE FEDERAL DA PARAÍBA
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
UNIVERSIDADE ESTADUAL DE CIÊNCIAS DA SAÚDE DE ALAGOAS
PROGRAMA ASSOCIADO DE PÓS-GRADUAÇÃO EM FONOAUDIOLOGIA**

ATA DE DEFESA DE DISSERTAÇÃO

Aos treze dias do mês de dezembro de 2021 (13/12/2021), às 14:00 horas, realizou-se na plataforma de videoconferência Zoom meet, por meio do link <https://us02web.zoom.us/j/82001530119?pwd=bGJHaVE1dWxhSHpHOVVEVEdjA3SElidz09> a sessão pública de defesa de dissertação intitulada “**BANCO VOZES BRASILEIRO NAS VARIAÇÕES DAS EMOÇÕES (EMOVOX-BR): ESTRUTURAÇÃO E VALIDAÇÃO**”, apresentada pela mestrand **Héryka Maria Oliveira Lima**, que concluiu os créditos para obtenção do título de MESTRE EM FONOAUDIOLOGIA, área de concentração Aspectos Funcionais e Reabilitação em Fonoaudiologia, segundo encaminhamento do Prof. Dr. Leandro de Araújo Pernambuco, coordenador do Programa Associado de Pós-Graduação em Fonoaudiologia da UFPB/UFRN/UNCISAL e segundo registros constantes nos arquivos da Secretaria da Coordenação do Programa. A Profa. Dra. Anna Alice Figueiredo de Almeida Queiroz, na qualidade de orientadora presidiu a Banca Examinadora da qual fizeram parte o Prof. Dr. Leonardo Wanderley Lopes (Examinador Interno/UFPB) e o Prof. Dr. Nelson Torro Alves (Examinador Externo/UFPB). Dando início aos trabalhos, a senhora presidente Profa. Dra. . Anna Alice Figueiredo de Almeida Queiroz convidou os membros da banca examinadora para compor a mesa. Em seguida, foi concedida a palavra à mestrand para apresentar uma síntese de sua dissertação. Posteriormente, a mestrand foi arguida pelos membros da banca examinadora. Encerrando os trabalhos de arguição, os examinadores deram o parecer final sobre a dissertação, ao qual foi atribuído o conceito de APROVADO. Proclamado o resultado pela Profa. Dra. Anna Alice Figueiredo de Almeida Queiroz, presidente da banca examinadora, foram encerrados os trabalhos e, para constar a presente ata foi lavrada e assinada por todos os membros da banca examinadora.

Link para gravação da sessão de defesa de dissertação:
https://drive.google.com/drive/folders/10vZdjWhLLJIfNpdW_6AatYe_M2nRfWWm?usp=sharing

João Pessoa/Natal/Maceió, 13 de dezembro de 2021

Profa. Dra. Anna Alice Figueiredo de Almeida Queiroz
(Presidente da Banca Examinadora)

Prof. Dr. Leonardo Wanderley Lopes
(Membro Interno)

Prof. Dr. Nelson Torro Alves
(Membro Externo – UFPB)



UNIVERSIDADE FEDERAL DA PARAÍBA
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
PROGRAMA ASSOCIADO DE PÓS-GRADUAÇÃO EM FONOAUDIOLOGIA

BANCO DE VOZES BRASILEIRO NAS VARIAÇÕES DAS EMOÇÕES (EMOVOX-BR): ELABORAÇÃO E VALIDAÇÃO

Héryka Maria Oliveira Lima

Dissertação apresentada à banca examinadora, constituída pelos seguintes professores:

Profa. Dra. Anna Alice Almeida (Orientadora – UFPB)

Prof. Dr. Leonardo Wanderley Lopes (Avaliador Interno – UFPB)

Prof. Dr. Nelson Torro (Avaliador Externo-UFPB)

13 de Dezembro 2021

AGRADECIMENTOS

A Deus, pelo dom da vida e saúde em tempos difíceis pelas quais passamos, e por me permitir ultrapassar todos os obstáculos encontrados ao longo da construção e realização deste trabalho. Aos meus pais e irmãos, por me ensinar o valor do esforço para conquistar meus objetivos e nunca desistir dos sonhos, sendo inspiração de resiliência e perseverança.

A minha querida orientadora, Dra. Anna Alice Almeida, por todos os conselhos, palavras de carinho, disponibilidade e por desenvolver minhas potencialidades, acreditando que juntas poderíamos dar forma a este trabalho, conduzindo a minha orientação sempre gentil e paciente.

Ao meu parceiro de vida Romário Ramos, por me apoiar emocionalmente nas adversidades passadas, ser compreensivo e paciente e por estar ao meu lado durante todo o percurso de desenvolvimento desta dissertação.

Às brilhantes alunas da Iniciação científica, Giglyene Monteiro, Andreia Morais, Soeme Ferreira que foram parceiras na construção deste trabalho, auxiliando e sempre que necessário e se disponibilizando para prestar apoio. A brilhante Larissa pelo acolhimento, sensibilidade e paciência durante todo o desenvolvimento dessa dissertação com contribuições científicas essenciais. A querida Priscila Oliveira, pela inspiração, carinho, e por ter dado um grande incentivo para que pudesse adentrar no caminho da ciência.

A banca de avaliação do presente estudo, por todo incentivo, dedicação e contribuições essenciais para a finalização desta dissertação.

Ao LIEV, por ter acolhido meus objetivos e ter me desenvolvido como pesquisadora, proporcionando experiências únicas e discussões valiosas durante todo percurso acadêmico.

Por fim, agradeço a todos que de alguma forma contribuíram e torceram para a conclusão desse objetivo em minha vida.

A todos, meu sincero obrigado!

RESUMO

A presente dissertação ampliou estudos com a temática relacionada à voz e emoções a partir da criação e validação de um banco de vozes com variações emocionais de indivíduos provenientes da população Brasileira. Esta dissertação é dividida em dois artigos: 1) É um estudo observacional e transversal que teve o objetivo validar a construção de um banco de vozes e elencar os achados da avaliação perceptivoauditiva de juízes experts em voz e o 2) É um estudo transversal, descritivo e quantitativo que se propôs analisar parâmetros acústicos de um Banco Brasileiro de Vozes nas variadas emoções (EMOVOX-BR). O primeiro estudo contou com 182 áudios produzidos por atores ou estudantes de Artes Cênicas, que obtiveram melhor qualidade de relação sinal ruído (SNR). Esses passaram por avaliação perceptivoauditiva pelos juízes fonoaudiólogos com expertise em voz. Conclui-se que os fonoaudiólogos conseguiram identificar as emoções, valência e potência, além de perceberem as modificações de voz e fala nas variadas emoções. No segundo estudo foram selecionados os 39 áudios que obtiveram melhores resultados da análise do grau de confiabilidade intra-áudios e concordância inter-juízes. Foram extraídas as medidas frequência fundamental (F0) média, desvio, mínimo e máximo; *jitter*; *shimmer*; proporção harmônico-ruído (PHR); intensidade; média e máximo de *extracion glotal noise* (GNE). Os resultados obtidos permitem identificar características de variações importantes nos parâmetros acústicos na expressão das emoções por meio da voz.

Palavras-chave: Fonoaudiologia, voz, emoções, comportamento, acústica.

ABSTRACT

This dissertation expanded studies on the theme related to voice and emotions from the creation and validation of a voice bank with emotional variations of individuals from the Brazilian population. This dissertation is divided into two articles: 1) It is an observational and cross-sectional study that aimed to validate the construction of a voice bank and to list the findings of the perceptive evaluation of expert judges in voice and 2) It is a cross-sectional study, descriptive and quantitative that proposed to analyze acoustic parameters of a Brazilian Bank of Voices in the varied emotions (EMOVOX-BR). The first study had 182 audios produced by actors or students of Performing Arts, who obtained better quality of signal noise ratio (SNR). These underwent auditory perceptual evaluation by speech therapists with voice expertise. It is concluded that the speech therapists were able to identify the emotions, valence and power, besides realizing the voice and speech modifications in the varied emotions. In the second study were selected the 39 audios that obtained better results of the analysis of the degree of intra-audios reliability and inter-judges agreement. The fundamental frequency (F0), mean, deviation, minimum and maximum measurements were extracted; *jitter*; *Shimmer*; harmonic-to-noise ratio (PHR); intensity; mean and *maximum glottal extraction* (GNE). The results obtained allow identifying characteristics of important variations in acoustic parameters in the expression of emotions through voice.

Key Words: Speech therapy, voice, emotions, behavior, acoustics.



UNIVERSIDADE FEDERAL DA PARAÍBA

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
PROGRAMA ASSOCIADO DE PÓS-GRADUAÇÃO EM FONOAUDIOLOGIA

SUMÁRIO

1. INTRODUÇÃO GERAL	10
2. ARTIGO 1-.....	14
2.1 Validação e taxa de acerto do EMOVOX-BR.....	14
2.1.1 Introdução.....	16
2.1.2 Métodos	17
2.1.3 Resultados	23
2.1.4 Discussão	26
2.1.5 Conclusão.....	28
2.1.6 Referências.....	29
Anexo I	
3. ARTIGO 2-.....	32
3.1 Extração de medidas acústicas e perceptuais na variação das emoções	32
3.1.1 Introdução.....	34
3.1.2 Métodos	36
3.1.3 Resultados	38
3.1.4 Discussão	40
3.1.5 Conclusão.....	42
3.1.6 Referências.....	43
Anexo I	
4. CONSIDERAÇÕES FINAIS.....	47
REFERÊNCIAS.....	48



1. INTRODUÇÃO GERAL

A voz vem sendo ao longo dos anos amplamente estudada quanto a sua relação com as emoções, sendo apontada como instrumento indicador do estado de humor de um indivíduo. Em convergência, por meio do estado de humor do indivíduo pode ser possível perceber as variações emocionais com base na escuta de sua voz (Brasil et al, 2017; Sundberg et al, 2015; Almeida et al, 2014; Seifert; Burkhardt, 2005; Trajano 2016; Almeida et al, 2011; Costa et al, 2013; Gomes et al, 2018).

A expressividade das emoções por meio da voz é de grande relevância social, como também participa do processo evolutivo da humanidade. No ato de se comunicar, as emoções são transmitidas de forma consciente ou inconsciente, fato que permite ao ouvinte perceber as emoções do interlocutor através da fala. A transmissão dessa mensagem poderá causar efeitos nas emoções e comportamentos do indivíduo que irá receber, dependendo da sua sensibilidade e precisão no reconhecimento. (Adriano; Arriaga, 2016; Vieira, 2018).

As emoções podem ser expressas e vinculadas a diferentes parâmetros, como, parâmetros acústicos vocais ou por meio de expressões faciais, a literatura sempre buscou um meio universal de classificações quanto às emoções, como o exemplo de alegria, tristeza e medo (Almeida, et al 2014). Foram universalmente reconhecidas na face pelos seres humanos seis emoções básicas (alegria, medo, raiva, tristeza, nojo, surpresa), após estudo realizados em diferentes culturas. (Ekman, 1992).

A interferência de aspectos emocionais na voz torna o discurso adaptável às diferentes situações vivenciadas, juntamente com a variação de prosódia e intensidade de fala (Sundberg et al, 2015). As emoções acarretam modificações momentâneas na anatomofisiologia do trato vocal que irá interferir no controle da respiração, no posicionamento vertical da laringe, no relaxamento relativo das pregas vocais no posicionamento e no relaxamento dos músculos da faringe e da língua (Behlau et al, 2008). Essas modificações trarão efeitos na pressão supraglótica e levará a alteração de intensidade e até da frequência de fonação, além disso também pode alterar a velocidade da fala, a coordenação pneumofonoarticulatória, a inteligibilidade de fala e a fluência com a presença de hesitações e repetições de palavras (Sundberg et al, 2015; Barbosa; Friedman, 2007). Na análise acústica das vozes é possível

perceber as modificações ocorridas e elencar os dados para caracterizar o sinal vocal, como os dados de frequência fundamental, jitter, shimmer e proporção harmônico-ruído entre outros. (Fellipe et, al 2006).

Atualmente o conhecimento sobre a variação das emoções e a construção de um banco de vozes composto por esses segmentos, vem sendo estudado, existem alguns bancos de vozes que foram construídos com a variação as emoções em outras línguas, são eles a base EMO-DB, que foi desenvolvida na Universidade Técnica de Berlin, na Alemanha. Participaram 40 atores das gravações de amostras vocais, em alemão, de 6 emoções: raiva, tédio, nojo, medo, felicidade, tristeza e neutro (Burkhardt et al,2005). A base IEMOCAP foi desenvolvida na Universidade do Sul da Califórnia, nos Estados Unidos. Os sinais acústicos foram realizados a partir das interações dois a dois, com 10 atores (5 homens e 5 mulheres), onde 7 eram atores profissionais e 3 eram alunos sênior no Departamento de Drama. As gravações foram no idioma inglês, e as emoções gravadas foram: raiva, felicidade, tristeza e emissão neutra (Busso et al, 2008).

A base SEMAINE é audiovisual e foi desenvolvida a partir de uma cooperação de Universidades da Inglaterra, da Holanda e da Alemanha. Participaram da coleta 150 estudantes de graduação e pós-graduação de oito diferentes países, todos padronizados no idioma inglês. As gravações foram feitas a partir de interações realizadas dois a dois, onde o controle ficava sempre a cargo de um dos dois. As emoções foram divididas em 27 estilos, dentre raiva e tristeza, além de outros comportamentos (McKeown et al, 2012).

A base RECOLA foi desenvolvida na Suíça, e utilizou-se o idioma francês, apesar de terem participantes nativos de outros idiomas. Um total de 46 participantes foram submetidos a interações entre os dois. Antes de haver a interação entre os participantes, eles foram submetidos a um questionário de autoavaliação emocional conhecido como Self-Assessment Manikin (SAM), que 10 está relacionado com a valência das emoções. Os aplicadores do questionário decidiram quais participantes iriam ser induzidos com humor positivo ou negativo, de acordo com o SAM. Nesse, capturaram sinais acústicos e biológicos, onde posteriormente extraíram dados dos sinais acústicos quanto à valência e ativação (Ringeval et al,2013).

Esses bancos de vozes desenvolvidos até então foram utilizados para analisar variações acústicas emocionais, pouco se tem sobre outros tipos de medidas, como por exemplo: medidas prosódicas, diferentes tarefas de fala, avaliação perceptiva auditiva por parte de juízes, análise acústicas e prosódicas de falantes nativos brasileiros das variações

emocionais. De toda forma, ainda não há um consenso a respeito de um atributo acústico, de qualidade vocal ou prosódicos que sinalize nas variadas emoções.

A construção de um banco de vozes que contemple diferentes tarefas de fala nas variações emocionais, dados sobre avaliação perceptivoauditiva por parte de juízes, e medidas acústicas de falantes do português Brasileiro, poderá contribuir com a definição de um atributo que represente de forma significativa informações relacionadas à expressão das emoções por meio da voz.

Baseando-se nesse cenário, essa dissertação buscou responder a seguinte pergunta: É possível identificar parâmetros vocais e de fala que sejam capazes de discriminar as emoções em falantes da população brasileira? A hipótese é que a partir da validação do banco de vozes em diferentes emoções e caracterização da análise acústica, seja possível encontrar parâmetros capazes de discriminar as emoções nos falantes do português brasileiro.

Portanto, a finalidade dessa dissertação foi elaborar e validar um banco de vozes com variações emocionais que contemplasse dados de voz e de fala de indivíduos provenientes da população brasileira. Para isto foram desenvolvidos dois artigos: o primeiro se configura como um estudo observacional e transversal uma vez que se propôs validar a construção do banco de vozes (EMOVOX-BR), já o segundo é um estudo transversal, descritivo e quantitativo pois se propôs analisar os parâmetros acústicos extraídos do EMOVOX-BR.

O primeiro artigo buscou responder as seguintes perguntas: Existem características vocais marcantes de acordo com a variação das emoções? Quais são as emoções mais facilmente identificadas por juízes fonoaudiólogos? A hipótese é que existam características vocais marcantes e essenciais para identificação das emoções por meio da escuta das vozes.

O segundo artigo se propõe a responder às seguintes perguntas: Existem parâmetros acústicos da voz que podem indicar as emoções simuladas em um banco de vozes? Quais os parâmetros acústicos indicam as emoções? A hipótese é que existem parâmetros acústicos capazes de diferenciar as variadas emoções.

Do ponto de vista para ciências da saúde, a dissertação busca contribuir na construção de conhecimento de características de voz e fala nas variadas emoções no português brasileiro, devido ao processo de elaboração e validação de um banco de vozes com variações emocionais inédito. Para a sociedade se propõe a auxiliar na construção de um sistema de reconhecimento por parte de ouvintes, uma vez que julgamentos pela voz, sejam acurados ou não, influenciam as interações sociais, a escolha de parceiros, líderes e até mesmo opções como consumidores (Klofstad et al, 2012; Tigue et al, 2012). Por fim, para a pesquisa e

produção de ciência, possui o objetivo de contribuir com levantamento de parâmetros possíveis para cada emoção e trazer novos achados científicos sobre o tema.

O objetivo desta dissertação é elaborar e validar um banco de vozes com variações emocionais que contemplem dados acústicos e de fala de indivíduos provenientes da população brasileira.



UNIVERSIDADE FEDERAL DA PARAÍBA
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
PROGRAMA ASSOCIADO DE PÓS-GRADUAÇÃO EM FONOAUDIOLOGIA

2. ARTIGO 1- SUGERE SUBMISSÃO AO LOGOPEDICS PHONIATRICS VOCALOGY

(ISSN: 1401-5439)

QUALIS A2 NA ÁREA 21 NO QUADRIÊNIO 2013-2016

ELABORAÇÃO E VALIDAÇÃO DO EMOVOX-BR

Héryka Maria Oliveira Lima ^{a*}, Anna Alice Almeida^b

^a Fonoaudióloga, Mestranda do Programa Associado de Pós-Graduação em Fonoaudiologia, Departamento de Ciências da Saúde, Universidade Federal da Paraíba, João Pessoa, Brasil; ^b Docente e pesquisadora do Departamento de Fonoaudiologia, Universidade Federal da Paraíba, João Pessoa, Brasil.

Resumo

Introdução: A voz desempenha uma importante função na expressão das emoções dos indivíduos, possibilita comunicar personalidade e estados emocionais, reconhecer características individuais na expressão das emoções por meio da escuta das vozes, pode influenciar na tomada de decisão de ações e comportamentos sociais. **Objetivo:** Elaborar e validar o Banco de Vozes Brasileiro nas Variações das Emoções (EMOVOX-BR). **Métodos:** Estudo de observacional e transversal. O *corpus* deste estudo foi constituído por 1.638 sinais sonoros produzidos por 26 atores ou alunos de Artes Cênicas brasileiros, gravado de forma remota por meio da plataforma Zoom Meetings, foram coletadas 3 tarefas de fala, nas variadas emoções (alegria, tristeza, medo, raiva, surpresa, nojo) e neutra. Na etapa de validação, a amostra foi composta por 10 fonoaudiólogas juízas com expertise na área de voz, com objetivo de realizar a análise perceptivouditiva das vozes selecionadas do EMOVOX-BR. **Resultados:** O EMOVOX-BR foi formado por 39 vozes, 15 vozes são do sexo feminino e 24 vozes do sexo masculino. Na segunda etapa, todas as emoções obtiveram uma alta taxa de acerto, superior a 70%, as emoções raiva, nojo e neutra foram as mais facilmente identificadas, já os parâmetros de voz e fala que sobressaíram na identificação das emoções foram o *pitch* e *loudness* segundo a avaliação das juízas. **Conclusão:** O EMOVOX-BR foi formado por 39 vozes que apresentam características marcantes, sinalizadas por juízas fonoaudiólogas após análise perceptivouditiva das vozes nas variadas emoções.

Palavras chaves: Voz, Emoções, Comportamento, Banco de Dados, Comunicação

Abstract

Introduction: The voice plays an important role in the expression of the emotions of individuals, it makes it possible to communicate personality and emotional states, to recognize individual characteristics in the expression of emotions by listening to voices, can influence the decision-making of social actions and behaviors. **Objective:** Develop and validate the Brazilian Voices Database on Variations of Emotions (EMOVOX-BR). **Methods:** Observational and cross-sectional study. The corpus of this study was composed of 182 audios produced by 26 actors or students of Brazilian Performing Arts, recorded remotely through the platform Zoom Meetings, were collected 3 speech tasks, in various emotions (joy, sadness, fear, anger, surprise, disgust) and neutral. In the validation stage, the sample was composed of 10 speech therapists with expertise in the voice area, in order to perform the perceptive analysis of the selected voices EMOVOX-BR. **Resultados:** The EMOVOX-BR was formed by 15 voices are female and 24 male voices. In the second stage, all emotions achieved a high hit rate, higher than 70%, emotions anger, disgust and neutral were the most easily identified, already the voice and speech parameters that stood out in the identification of emotions were the pitch and Loudness according to the evaluation of the judges. **Conclusão:** The EMOVOX-BR is composed of 39 audios of voices that present striking characteristics, signaled by speech-language judges after perceptive analysis of voices in various emotions.

Key Words: Voice, Emotions, Behavior, Database, Communication

Introdução

As emoções estão presentes no cotidiano dos seres humanos e podem repercutir diretamente em diferentes esferas, como por meio de reações fisiológicas, comportamentais e na comunicação. Mudanças do estado emocional resultam em um impacto direto na comunicação humana, sendo possível perceber as variações na escuta da voz.^[1,2]

A voz pode apresentar variações de intensidade, frequência e ritmo, de acordo com as emoções. Já as emoções interferem na produção da voz, sendo possível observar sua expressão por traços de personalidade, sentimentos, humor, dentre outros.^[3,4,5,6]

A relação entre voz e emoção, ao longo dos anos vem sendo amplamente estudada,^[7,8] pesquisas por características universais no reconhecimento das emoções, são reconhecidas na literatura, como na revisão sistemática realizada em diferentes culturas, em conclusão, propuseram 6 emoções básicas por serem expressa de forma semelhante nas diferentes culturas investigadas e possuírem configurações específicas comuns as diferentes culturas onde o estudo foi realizado.^[9,10]

Existem algumas iniciativas desenvolvidas com intuito de formar banco de vozes, para auxiliar no processo de construção e conhecimento acerca do comportamento vocal na expressão das emoções, destacam-se as bases internacionais: Berlin Database of Emotional Speech (EMO-DB)^[11], Interactive Emotional Dyadic Motion Capture (IEMOCAP)^[12], Sustained Emotionally colored Machine-human Interaction using Nonverbal Expression (SEMAINE)^[13] e Remote COLaborative and Affective interactions (RECOLA)^[14], essas são reconhecidas na literatura e detém informações acerca de variações acústicas emocionais.

A elaboração e validação de um banco de vozes com variações emocionais que detenham sobre dados de avaliação perceptivoauditiva por parte de juízes fonoaudiólogos e falantes do português Brasileiro, busca colaborar com a identificação de parâmetros de voz e fala específicas de cada estado emocional. Esses achados podem auxiliar na criação de padrões, que são importantes para o desenvolvimento de sistemas de interação homem - máquina na identificação de emoções, que poderá abarcar diversos tipos de mercado como *call center*, aplicativos que envolvem reconhecimento de voz, web filmes, comunicação móvel, entre outros.^[15]

O objetivo deste estudo é elaborar e validar o banco de vozes brasileiro nas variações das Emoções (EMOVOX-BR), com a finalidade de analisar se juízes fonoaudiólogos com expertise em voz identificam as emoções expressas nos áudios, e quais as características de voz e fala são marcantes na identificação das emoções.

Método

Esta é uma pesquisa observacional e transversal, avaliada e aprovada pelo Comitê de Ética em Pesquisa do Centro de Ciências da Saúde de uma instituição de ensino superior no Nordeste do Brasil, sob número 3.304.419/2018. (ANEXO 1).

Amostra

a) Elaboração do EMOVOX-BR

Todos os participantes da pesquisa deveriam seguir os seguintes critérios de elegibilidade: ausência de comorbidades que comprometessem a cognição, audição e comunicação que possa limitar a realização das tarefas solicitadas; não fazer uso de substâncias psicotrópicas; responder o google forms; ter acesso a internet, uso de microfone externo, e smartphone; ter realizado as gravações nas 6 variações das emoções e na emissão neutra. O *corpus* deste estudo foi constituído por 182 áudios produzidos por 26 atores ou alunos de Artes Cênicas brasileiros, majoritariamente atores profissionais, de ambos os sexos, com média de idade sendo 27 anos (DP=6,75).

A tabela 1 fornece dados de caracterização da amostra que compõem o banco de vozes nas variações emocionais.

Tabela 1. Caracterização da amostra do *corpus* do EMOVOX -BR

Variáveis	Frequência	Porcentagem
Sexo		
Feminino	11	42,3%
Masculino	15	57,6%
Estado civil		
Solteiro(a)	23	88,4%
Casado(a)	2	7,6%
Divorciado(a)	1	3,8%
Grau de Instrução		
Fundamental Incompleto	0	0%
Fundamental completo	0	0%
Ensino médio	2	7,6%
Ensino Superior incompleto	19	73%
Ensino Superior completo	4	15,3%
Pós-Graduação	1	3,8%
Profissão		
Atores profissionais	16	61,53%
Estudantes de artes cênicas	10	38,46%

Participação de companhia teatral		
Sim	13	50%
Não	13	50%
Tempo de atuação		
0-2 anos	8	30,7%
3-8 anos	6	23%
9-12 anos	9	34,6%
Maior que 12 anos	3	11,5%
Possui Smartphone		
Sim	26	100%
Não	0	0%

b) Validação do EMOVOX-BR

Para participar dessa etapa, os voluntários deveriam seguir os seguintes critérios de elegibilidade: ser fonoaudiólogo, possuir expertise na área de voz, não possuir alteração auditiva autorreferida e/ou diagnosticada e preencher o questionário hospedado no google forms, com dados sociodemográficos e avaliação perceptual das vozes.

Na etapa de validação, a amostra foi composta por 10 fonoaudiólogas juízas com expertise na área de voz (Tabela 2).

Tabela 2. Caracterização sociodemográfica e de formação dos juízes fonoaudiólogos

Variáveis	Frequência	Porcentagem
Sexo		
Feminino	100%	100%
Masculino	0	0%
Estado Civil		
Solteira	5	50%
Casada	4	40%
União estável	0	0%
Outros	1	10%
Renda Familiar		
Até 1 salário mínimo	0	0%
De 2 a 3 salários mínimos	1	10%
De 4 a 6 salários mínimos	3	30%
De 7 a 9 salários mínimos	1	10%
Mais de 9 salários mínimos	5	50%
Grau de Instrução		
Graduação	5	50%
Mestrado	0	0%
Doutorado	0	0%
Pós-Doutorado	5	50%
Tempo de Formação		
Menos de 1 anos	0	0%
De 1 a 5 anos	4	40%
De 6 a 10 anos	0	0%

Mais de 10 anos	6	60%
Possui Especialização na área de voz		
Sim	6	60%
Não	4	40%
Tempo de atuação na área da voz		
Menos de 1 anos	0	0%
De 1 a 5 anos	4	40%
De 6 a 10 anos	0	0%
Mais de 10 anos	6	60%
Possui alguma alteração auditiva		
Sim	0	0%
Não	10	100%

Materiais

a) Elaboração do EMOVOX-BR

As ferramentas utilizadas na primeira etapa da coleta foram: o questionário hospedado no *Google Forms*, onde buscou levantar dados sociodemográficos dos atores ou estudantes de Artes Cênicas, sendo esse composto por 12 itens que abordavam questões como: nome, idade, sexo, estado civil, grau de instrução, data e local de nascimento, endereço, e-mail, telefone, profissão e renda familiar, também foram coletados dados acerca da participação em alguma companhia teatral, o tempo e o período de trabalho, além de investigar se o voluntário possuía *smartphone* e/ou computador e o sistema operacional utilizado.

A Plataforma Zoom Meeting de chamada de vídeo, foi escolhida por sua praticidade e fácil acesso em sua versão gratuita, e por possuir segurança de ponta a ponta dos dados ^[16]. Ainda, foi utilizado aplicativo Audacity versão 3.0.2, com o uso desta ferramenta todos os sinais foram salvos no formato “wav” para que se mantivesse a melhor qualidade, sem perdas, no computador do pesquisador. Foram coletadas três tarefas de fala: vogal sustentada /e/; fala automática com contagem de números de 1 a 10; e fala semi-espontânea com uma frase que compõem o CAPE-V “Olha lá o avião azul” ^[17, 18], na variação das emoções.

b) Validação do EMOVOX-BR

A segunda etapa de coleta envolveu o uso do formulário *online*, hospedado no *Google Forms*, onde continham questionamentos sobre os dados sociodemográficos dos juízes fonoaudiólogos com expertise na área de voz, sendo esse composto por 12 itens que

abordavam o nome, idade, sexo, estado civil, data e local de nascimento, endereço, e-mail, telefone, renda familiar, grau de instrução, também foram coletados dados acerca do tempo de formação, se possuíam especialização na área de voz e alguma alteração auditiva.

Em uma segunda seção do próprio formulário, as fonoaudiólogas foram orientadas a ouvir 200 áudios. Desses, 18 áudios foram apresentados de forma repetida, esse valor representa 10% da amostra e foi apresentado de forma randomizada, sendo solicitado após a escuta dos 200 áudios assinalar no questionário a emoção que foi identificada, a intensidade que esta emoção expressava, valência (positiva, negativa ou neutra), além de assinalar dados referentes a aspectos de voz e fala que identificavam cada uma das emoções com base na escuta dos áudios hospedado no formulário.

Procedimentos de coleta de dados

a) Elaboração do EMOVOX-BR

Inicialmente o projeto foi divulgado por meio de redes sociais, os voluntários que demonstraram desejo em participar da pesquisa foram informados quanto aos objetivos da pesquisa, receberam instruções sobre as tarefas de fala e simulação das emoções previamente, além de ler e concordar com o Termo de Consentimento Livre e Esclarecido (TCLE). O TCLE ainda foi encaminhado por e-mail com a segunda via assinada pelo pesquisador responsável.

Os voluntários responderam um questionário hospedado no Google Forms. Esse coletou dados sociodemográficos dos atores e estudantes de Artes Cênicas. Após essa coleta inicial, os voluntários recebiam um tutorial com roteiro e procedimentos de gravação e em seguida realizavam o agendamento para a coleta da voz online dos voluntários simulando as emoções e a emissão neutra. Foram coletadas três tarefas de fala distintas: 1) a emissão da vogal /ε/ sustentada, 2) fala automática a partir da contagem de 1 a 10 e 3) fala semi-espontânea por meio da frase “Olha lá o avião azul.”, que compõe o *Consensus Auditory Perceptual Evaluation of Voice* (CAPE-V) ^[17], validado para o português Brasileiro ^[18].

Diante de estudos realizados anteriormente ^[19,20], onde buscou analisar a qualidade de diferentes tarefas de fala nas variações emocionais, como também investigar quais as diferenças entre a qualidade do sinal coletado de forma remota com entrada line in e pela plataforma zoom meeting, também foram elencados trabalhos, que objetivavam analisar as vozes coletadas via smartphone possuíam relação sinal ruído diferente dentre as vozes coletas com o uso de microfone e sem o microfone externo. Os resultados desses estudos apontaram

que a tarefa de fala semi-espontânea e contagem são as que apresentam melhor qualidade da relação sinal ruído na coleta remota, a opção de entrada direta no aparelho para coleta da voz de forma line in sobressai quando comparada ao áudio gravado pela própria plataforma zoom meeting, na coleta por smartphone o uso do microfone não interfere na melhor qualidade de sinal, e o smartphone obteve resultados como uma opção segura e eficaz para a coleta remota da voz. [18,19,20]

Portanto, optou-se em selecionar os áudios com a tarefa de fala semi-espontânea por meio da frase do CAPE-V, que obtiveram a melhor qualidade da relação sinal ruído, coletada via smartphone sem microfone externo, por meio de uma entrada do som line in. Portanto, foram selecionados 182 sinais de áudio para seguir a etapa de análise perceptivoauditivo a ser realizada pelas juízas fonoaudiólogas, os áudios foram submetidos à análise de relação sinal-ruído, a signal to noise ratio – SNR busca saber qual a diferença de sinal é necessário para considerar o áudio dentro do padrão com valor de referência igual ou superior a 30dB de acordo com o padrão preconizado pela literatura [16,21].

b) Validação do EMOVOX-BR

Nessa etapa, buscou-se coletar informações sobre a avaliação perceptivoauditiva realizada pelas juízas como também a avaliação dos parâmetros de voz e fala nas variadas emoções. As juízas fonoaudiólogas obtiveram acesso ao formulário hospedado no google forms, esse foi subdividido em duas sessões, inicialmente buscou coletar dados sociodemográficos e a segunda sessão foi composta com as vozes dos atores simulando as variadas emoções.

As juízas escutavam os áudios, analisavam e indicava a emoção correspondente àquele áudio, a intensidade da emoção (o quanto a emoção transmitida pelo áudio), valência (emoção positiva ou negativa), e quais parâmetros de voz e fala foram decisivos para a identificar a emoção daquele áudio (*pitch*, *loudness*, articulação, velocidade de fala, incoordenação pneumofonoarticulatória, fluência, qualidade vocal). Buscou-se investigar se a qualidade do áudio influenciou na análise perceptual das vozes segundo as juízas.

Avaliou-se o trecho do banco EMOVOX-BR, cerca de 182 vozes, com a tarefa de fala “olha lá o avião azul”. Dessas 182 vozes, visando proporcionar melhor avaliação foram adicionados repetição de 10% da amostra dos áudios de forma randomizada por meio da planilha eletrônica no Microsoft Office Excel, para a análise da confiabilidade interna. Assim, para a avaliação com as juízas foram utilizadas 200 vozes, com o objetivo de entender quais

vozes eram as mais confiáveis como também a concordância inter juízes

Análise dos dados

Foi realizada análise estatística descritiva com tabela de frequência e porcentagens e demais medidas descritivas por meio das respostas dos voluntários e juízes, além das medidas de média e desvio padrão, a depender do tipo de variável para verificar as características sociodemográficas dos indivíduos. Na sequência, foram empregadas técnicas estatísticas inferenciais.

Na estatística inferencial, foram consideradas as variáveis dependentes: emoções, já como variáveis independentes foram consideradas: valência, potência e parâmetros de voz e fala. Foi realizado a análise do grau de confiabilidade intra-áudios e concordância Inter-juízes, o teste de concordância de Kappa, no qual foi comparado o número de acertos das emoções, ou seja, quantas vezes as juízas assinalaram a emoção real simulada por atores por meio da avaliação das vozes. Com vistas a analisar a confiabilidade para avaliar a concordância intra e inter-juízes, foi utilizado o Coeficiente de Correlação Intraclasse (CCI) e o Kappa nos julgamentos das amostras vocais, de modo a fornecer indicações das amostras mais representativas de cada emoção.

O teste Qui-quadrado para verificar a associação entre as variáveis. Foi utilizado ainda o teste de correlação de Pearson, que possui o objetivo de medir a relação estatística entre duas variáveis na análise acústica, bem como a análise de variância (ANOVA) para fazer comparações entre os grupos e por fim o teste post hoc sendo esse realizado após os resultados do ANOVA, para identificar quais os pares, no caso as emoções se diferem dos grupos significantes.

Todas as análises foram realizadas por meio do software SPSS e utilizou-se o nível de significância de 5%.

Resultados

a) Elaboração do EMOVOX-BR

O EMOVOX-BR está estruturado por 39 áudios, desses, 25 áudios são de vozes masculinas e 15 áudios de vozes femininas (Table 3).

Table 3. Descrição dos áudios selecionados para formação o EMOVOX-BR

Variáveis	Frequência	Porcentagem
Sexo		
Masculino	24	61,5%
Feminino	15	38,4%
Emoção		
Surpresa	8	20,5%
Tristeza	7	17,9%
Raiva	7	17,9%
Neutra	7	17,9%
Medo	5	12,8%
Alegria	3	7,6%
Nojo	2	5,4%

Foi realizado a análise de confiabilidade do KAPPA com os 39 áudios que compõem o EMOVOX-BR, é possível observar que os valores intra áudios estão todos acima de 0,7 isto é, os áudios selecionados apresentam resultados confiáveis (Tabela 4).

Tabela 4. Descrição dos áudios e valor da confiabilidade no KAPPA

Variáveis	Sexo	Emoção	Valor de concordância KAPPA
Áudio 05	Masculino	Surpresa	0,855
Áudio 07	Masculino	Neutra	0,799
Áudio11	Masculino	Raiva	0,711
Áudio 12	Masculino	Surpresa	0,808
Áudio 14	Masculino	Neutra	0,799
Áudio15	Masculino	Alegria	1,0
Audio 18	Masculino	Raiva	1,0
Áudio 21	Masculino	Neutra	1,0
Áudio26	Feminino	Surpresa	0,714
Áudio 28	Feminino	Neutra	0,835
Áudio 30	Masculino	Medo	0,855
Áudio 33	Masculino	Surpresa	0,753
Áudio38	Feminino	Tristeza	1,0
Áudio45	Masculino	Tristeza	0,879
Áudio 52	Masculino	Tristeza	1,0
Áudio 54	Masculino	Surpresa	0,902
Áudio 56	Masculino	Neutra	0,822
Áudio 80	Masculino	Tristeza	0,855

Áudio 81	Masculino	Raiva	0,899
Áudio 82	Masculino	Surpresa	0,783
Áudio 105	Masculino	Neutra	0,877
Áudio 121	Feminino	Medo	1,0
Áudio123	Feminino	Raiva	0,799
Áudio127	Masculino	Alegria	0,835
Áudio 128	Masculino	Medo	0,893
Áudio 132	Masculino	Nojo	0,711
Áudio 137	Masculino	Raiva	0,714
			0,943
Áudio138	Masculino	Surpresa	0,783
Áudio139	Masculino	Nojo	1,0
Áudio 142	Feminino	Medo	0,713
Áudio 149	Feminino	Medo	0,855
Áudio 150	Feminino	Tristeza	1,0
Áudio 151	Feminino	Raiva	0,799
Áudio 154	Feminino	Neutra	0,825
Áudio 157	Feminino	Tristeza	1,0
Áudio 165	Feminino	Raiva	1,0
Áudio 178	Feminino	Tristeza	0,714
Áudio180	Feminino	Surpresa	0,808
Áudio 141	Feminino	Alegria	0,771

Quanto a concordância inter juízes, ou seja, a concordância entre todas as juízas que realizaram a análise perceptivoauditiva o valor obtido do KAPPA é de 0,711.

b) Validação do EMOVOX-BR

A tabela 5 apresenta o percentual de acerto dos áudios avaliados pelas juízas nas variadas emoções. Quando comparado o acerto em relação aos áudios, é possível observar que todas as emoções obtiveram percentual de acerto superior a 70%, configurando que há um alto percentual de acerto da emoção quando realizada a escuta dos áudios pelas juízas, indicando que, o trecho avaliado que compõem o banco EMOVOX-BR, expressa as emoções segundo a avaliação dos juízes.

Tabela 5. Percentual de acerto das variadas emoções por juízas Fonoaudiólogas.

Variáveis Emoção	Frequências	Percentual de acerto	P-Valor
Neutra	7	84,4%	0,0444
Raiva	7	82,6%	0,0444
Nojo	2	75%	0,0444
Surpresa	7	73,8%	0,0444
Alegria	3	73,3%	0,0444
Tristeza	7	72,5%	0,0444
Medo	5	71,3%	0,0444

Teste Qui-quadrado de Pearson; significância $p < 0,05$

A Tabela 6 mostra a identificação de valência atribuída pelas juízas para as emoções avaliadas. Observou-se que a percepção da valência quanto à classificação das emoções nas vozes avaliadas por juízas fonoaudiólogas foi definida como: alegria positiva, medo negativa, tristeza negativa, raiva negativa, surpresa positiva, nojo negativa e a emissão neutra como de valência neutra.

Tabela 6. Descrição da identificação dos parâmetros de valência na variação das emoções na emissão de vozes acordo com as impressões de juízas fonoaudiólogas.

Valência	Emoção												P-valor		
	Alegria		Medo		Tristeza		Raiva		Surpresa		Nojo			Neutra	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	
Positiva	27	90%	5	8,3%	1	1,3%	2	2,9%	70	85,7%	4	4%	2	2,9%	0,0001
Negativa	3	10%	51	85%	61	76,3%	63	90%	9	11,3%	16	16%	12	17,1%	0,0001
Neutra	0	0%	4	6,7%	18	22,5%	5	7,1%	1	1,3%	0	0%	56	80%	0,0001

Teste Qui-quadrado de Pearson; significância $p < 0,05$

A Tabela 7 apresenta a identificação dos parâmetros de voz e fala assinalados como mais importantes na caracterização das emoções apresentada pelas juízas na avaliação perceptivoauditiva.

Tabela 7. Identificação dos parâmetros de voz e fala na caracterização das emoções de acordo com as impressões de juízas fonoaudiólogas.

Parâmetros de voz e Fala	Emoção												P-valor		
	Alegria		Medo		Tristeza		Raiva		Surpresa		Nojo			Neutra	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	
Pitch	21	70%	30	51,7%	49	61,3%	40	57,1%	62	77,5%	13	65%	36	51,4%	0,0039
Loudness	11	36,7%	28	46,7%	28	35%	35	50%	31	38,8%	5	25%	17	24,3%	0,0035
Articulação	4	13,3%	4	6,7%	3	3,8%	17	24,3%	10	12,5%	4	20%	5	7,1%	0,0002
Velocidade F	11	36,7%	18	30%	33	41,3%	33	47,1%	28	36,3%	70	35%	24	34,3%	0,0045

Legenda: Velocidade F= velocidade de fala;. Teste Qui-quadrado de Pearson; significância $p < 0,05$

Observa-se que no presente estudo o *pitch* foi um parâmetro essencial para identificar todas as emoções básicas: alegria, medo, tristeza, raiva, surpresa e nojo. A *loudness* foi importante para reconhecer a emoção raiva. Já articulação, velocidade de fala, ICPFA, qualidade vocal e fluência não foram parâmetros decisivos para a determinar a classificação das emoções apresentadas, todas com resultado inferior a 50%.

Discussão

A construção do pioneiro banco de vozes brasileiro nas variações das emoções EMOVOX-BR é composto por 39 áudios, sendo eles 24 com vozes masculinas e 15 áudios com vozes femininas, pertencem a emoção surpresa oito áudios, as emoções raiva, tristeza e neutra possuem sete áudios cada, já a emoção medo são cinco áudios, emoção alegria três e por fim a emoção nojo contendo dois áudios. Baseando-se nos estudos científicos, onde as seis emoções básicas são reconhecidas por “Big-Six”, foi-se acrescentando a eles para compor o banco a neutra que diz respeito ao sinal de áudio que não tem predomínio de alguma das emoções [16,17,21].

De acordo com estudos anteriores sobre detecção de baixa e alta ansiedade na análise de vozes com leigos e profissionais generalista obtiveram taxa de acerto em torno de 50% [19,20]. O presente estudo realizado com o grupo de profissionais com expertise na área de voz, apresentou alto percentual de acerto das seis emoções básicas, com valores de referência superiores a 70% em todas as emoções (alegria, medo, tristeza, raiva, surpresa, nojo e neutra), quando realizada a avaliação perceptivoauditiva das vozes por juízas. O alto percentual de acerto das emoções avaliadas pelas juízas, dos trechos que compõem o EMOVOX-BR, são representativas e capazes de expressar as emoções, segundo a avaliação das juízas. A emoção com maior percentual de acerto foi a emoção raiva e a neutra e a com menor percentual de acerto a emoção medo. Esse achado é confirmado pela literatura, onde a emoção raiva gera um impacto maior na identificação das emoções do interlocutor, para sua produção é empregado uma maior energia, e está relacionada a alterações de posicionamento de laringe, velocidade de fala e intensidade. [16, 17, 19]

Atualmente existem cerca de 14 bancos de vozes mundiais, desses apenas seis são de acesso aberto à comunidade científica, ainda se faz importante elencar que grande parte desses bancos de vozes foram realizados com adultos e dois são bancos de vozes infantis, sendo um voltado à variação de estresse. São esses a bases Berlin Database of Emotional Speech (EMO-DB)^[11], Interactive Emotional Dyadic Motion Capture (IEMOCAP) ^[12], Sustained Emotionally colored Machine-human Interaction using Nonverbal Expression (SEMAINE) ^[13] e Remote Collaborative and Affective interactions (RECOLA)^[14], por apresentar estrutura de construção com certa similaridade como: participantes, acesso a base e as emoções analisadas. Essas quatro bases internacionalmente reconhecidas foram até então desenvolvidas no intuito de analisar apenas as variações acústicas.

A valência das emoções é o aspecto que visa identificar o quão positiva ou negativa é a emoção. Nesse estudo, utilizou-se também do aspecto de valência neutra, ou seja, sem identificação das emoções na emissão, as juízas avaliaram as emoções de valência positiva (alegria e surpresa), valência negativa (medo, raiva, tristeza e nojo) e de valência neutra (neutra). A literatura apresenta achados similares, modificações de emoções e valências fazem parte da expressividade, a manutenção e repetição de padrões constantes como os de valências negativas podem causar alterações como ressecamento da mucosa oral interferindo na ação de fala, afetar o controle da musculatura oral e articulação dos fonemas, e pode provocar temores [7,24]. Assim, o estado emocional provoca alterações na atividade comunicativa e não apenas na voz [21,24].

Alguns parâmetros identificados na voz e fala podem ser entendidos como um conjunto de fenômenos suprasegmentais, são eles a velocidade de fala, ritmo no aspecto temporal, à organização melódica (acento, melodia, entonação) e à intensidade (volume, força) presentes na fala. [25,26]. A identificação dos parâmetros de voz e fala são importantes na caracterização das emoções na emissão dos áudios de acordo com as juízas fonoaudiólogas, a partir da avaliação perceptivoauditiva, verificou-se que o parâmetro *pitch* foi o mais presente para identificar as emoções: alegria, medo, tristeza, raiva, surpresa, nojo e neutra. Já o parâmetro *loudness* esteve presente na emoção raiva. Os demais parâmetros como, articulação, velocidade de fala, fluência, qualidade vocal, incoordenação pneumofonoarticulatória (ICPFA) não obtiveram dados suficientes para serem considerados como critério classificador das emoções.

De forma geral, encontra-se que o *pitch* e *loudness* se reafirmaram como parâmetros importantes para identificar as emoções por meio da voz [7,8,9,27]. Estudos apontam que a emoção alegria, é expressa a partir do *pitch* agudo, *loudness* forte e pausas curtas, na emoção raiva o *pitch* agudo, *loudness* fraca nos homens e forte nas mulheres e velocidade de fala aumentada para mulheres, na emoção nojo, representa-se com o *pitch* é grave, *loudness* fraca e velocidade de fala reduzida. [2,16, 27,28]

Portanto, juízas experientes puderam identificar as emoções em falantes nativos do português brasileiro e perceber características comuns nas variações emocionais, por meio da avaliação perceptivoauditiva. Esses achados contribuem a comunidade científica uma vez que possuem áudios testados e validados com alta confiabilidade, e difere das demais bases de dados, uma vez que é uma base composta por áudios em português brasileiro.

Conclusão

Foi elaborado o Banco de Vozes Brasileiro nas variações das emoções (EMOVOX-BR) com a composição de 39 áudios, representando as emoções alegria, medo, tristeza, raiva, surpresa, nojo e neutra. O EMOVOX-BR foi construído remotamente, com participação de atores na simulação das emoções e foram mantidos os áudios com qualidade satisfatória de relação sinal-ruído.

Na etapa de validação, os 39 áudios hospedados no banco EMOVOX-BR possuem alta confiabilidade quanto à qualidade dos áudios, identificação e valência das emoções. Os juízes fonoaudiólogos perceberam modificações de voz e fala nas variadas emoções com boa concordância inter juiz. O parâmetro *pitch* se sobressaiu para identificação de todas as emoções, e a *loudness* identificou a emoção raiva. Portanto, vê-se que as vozes variam e apresentam características marcantes de acordo com as emoções que podem ser identificadas perceptualmente por juízes treinados.

Referências

1. Behlau M. *Voz: O livro do especialista*. Vol. I. Rio de Janeiro: Editora Revinter, 2008.
2. Sundberg, j. *A ciência da voz*. Editora da universidade de são paulo, 2015.
3. Silva, EF. A voz dentro da relação psíquico-orgânica: estudo sobre a influência das emoções na voz do ator. *Rev Cient /FAP*, vol.4, n.1, p.1-19, 2009.
4. Costa, d. B.; lopes, l. W.; silva, e. G.; cunha, g. M. S.; almeida, l. N. A.; almeida, a. A. F. Fatores de risco e emocionais na voz de professores com e sem queixas vocais. *Revista cefac [online]*, v. 15, n. 4, p. 1001-1010, 2013. Disponível em: <https://doi.org/10.1590/s1516-18462013000400030>.
5. Lopes LW; Silva IM; Sousa ESS; Silva ACF; Paiva MAA; Diniz EGR; Silva POC. Spectrographic classification of the vocal signal: relation with laryngeal diagnosis and auditory-perceptual analysis. *Audiol Commun Res*. 2020;25:e2194.
6. Almeida, a. A. F.; behlau, m.; leite, jr. Correlação entre ansiedade e performance comunicativa. *Rev soc bras fonoaudiol*, v. 16, n.4, p.384-386, 2011.
7. Adriano, T.; Arriaga, P. Exaustão Emocional e Reconhecimento de Emoções na Face e Voz em Médicos. *Psicol saúde doenças, Lisboa*, v. 17, n. 1, p. 97-104, 2016.
8. Pereira, A L. The “voice colours”: Expressing Emotions Through the Singing Voice Timbre. *Cadernos de Saúde, Vol. 18, N.º 2*, p. 147-166, 2008.
9. Ekman P. *The Handbook of Cognition and Emotion*, ch. Basic Emotions. Wiley Online Library, pp. 45–60, 1999.
10. Ververidis D, Kotropoulos C. Emotional speech recognition: Resources, features, and methods. *Speech Communication, V.48*, p. 1162–1181, 2006.
11. Burkhardt F, Paeschke A, Rolfes M, Sendlmeier W, Weiss B. A Database of German Emotional Speech, in *Proc. INTERSPEECH*. Pp. 1517–1520, 2005.
12. Busso C, Bulut M, Lee CC, Kazemzadeh A, Mower E, Kim S, Chang JN, Lee S, Narayanan SS. IEMOCAP: Interactive Emotional Dyadic Motion Capture Database. *Language Resources and Evaluation.V.42(4)*, p. 335–59, 2008.
13. Mckeown G, Valstar M, Cowie R, Pantic M, Schroder M. The SEMAINE 24 Database: Annotated Multimodal Records of Emotionally Colored Conversations between a Person and a Limited Agent. *IEEE Transactions on Affective Computing. V.3(1)*, p. 5–17, 2012.
14. Ringeval F, Sonderegger A, Sauer J, Lalanne D. Introducing the recola multimodal corpus of remote collaborative and affective interactions. *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp. 1–8, 2013.
15. Abreu, M A L. *Desempenho de técnicas de Aprendizado de Máquina em Reconhecimento de Voz.[TCC]*. Santa Catarina/SC: instituto federal de santa catarina, 2019.
16. Florêncio, v. O.; almeida, A. A.; BALATA, P.; NASCIMENTO, S.; BROCKMANN-BAUSER, M.; LOPES, L. W. Differences and Reliability of Linear and Nonlinear Acoustic Measures as a

- Function of Vocal Intensity in Individuals With Voice Disorders [published online ahead of print, 2021 Jun 8]. *J Voice*. 2021; S0892-1997(21)00144-2. Doi:10.1016/j.jvoice.2021.04.011.
17. Behlau, m.; rocha, b.; englert, m.; madazio, g. Validation of the brazilian portuguese cape-v instrument—br cape-v for auditory-perceptual analysis. *J voice*, 2020 aug 15;s0892-1997(20)30257-5
 18. ASHA, American Speech-Language-Hearing Association. Consensus auditoryperceptual evaluation of voice (CAPE-V). Rockville: ASHA Special Interest Division 3, Voice and Voice Disorders; 2002.
 19. Morais, A S; Santos, S F. Julgamento perceptual a diferentes estados emocionais de pessoas com e sem problemas de voz na perspectiva de juízes leigos. *Iniciação Científica - Pró-Reitoria de Pesquisa*. Universidade Federal da Paraíba, João Pessoa, 2021.
 20. Monteiro, G F P. Será que o smartphone é uma boa estratégia de coleta de voz de forma remota. *Iniciação Científica - Pró-Reitoria de Pesquisa*. Universidade Federal da Paraíba, João Pessoa, 2021.
 21. Deliyiski, d. D.; shaw, h. S.; evans, m. K. Adverse effects of environmental noise on acoustic voice quality measurements. *J voice*, v. 19, n. 1, p. 15-28, 2005. Doi:10.1016/j.jvoice.2004.07.003.
 22. Bottalico, P.; Codino, J.; Cutiva, L. C.; Marks, K.; Nudelma, C. J.; Skeffing, J.; Shrivastav, R.; Jackson, M. C.; Hunter, E. J.; Rubin, A. D. Reproducibility of Voice Parameters: The Effect of Room Acoustics and Microphones. *J Voice*, v. 34, n. 3, p. 320-34, 2020. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0892199718304338>.
 23. Landis S, J. R.; Koch, G. G. A one way components of variance model for categorical data. *Biometrics*, v. 33, p. 671–679, 1977.
 24. ALVES, J. N. Influência da experiência do ouvinte e da tarefa de fala na avaliação perceptivo auditiva da qualidade vocal. 2019. 74 p. Dissertação (Mestrado em Linguística) - Universidade Federal da Paraíba, João Pessoa, 2019.
 25. Hirst, d.; di cristo, a. *Intonation systems*. Cambridge: cambridge university press, 1998.
 26. Antunes, l. B. O papel da prosódia na expressão de atitudes do locutor em questões. 306f. Tese (doutorado em letras: estudos linguísticos). Belo horizonte: ufmg, 2007.
 27. Dejonckere PH, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *Eur Arch Otorhinolaryngol*. 2001; 258:77-82.
 28. Vieira vjd. Análise de variações acústicas não estacionárias e seu efeito na detecção de múltiplas emoções e condições de estresse. [tese em engenharia elétrica]. Campina grande/pb: universidade federal de campina grande, 2018.

Anexo I

UFPB - CENTRO DE CIÊNCIAS
DA SAÚDE DA UNIVERSIDADE
FEDERAL DA PARAÍBA



Continuação do Parecer: 3.304.419

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1306092.pdf	03/04/2019 23:31:56		Aceito
Outros	Carta_resposta_CEP_UFPB.pdf	03/04/2019 23:31:25	Anna Alice Figueiredo de Almeida	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE2.pdf	03/04/2019 23:28:03	Anna Alice Figueiredo de Almeida	Aceito
Declaração de Pesquisadores	Atestado_MBehlau.pdf	03/04/2019 23:27:46	Anna Alice Figueiredo de Almeida	Aceito
Declaração de Instituição e Infraestrutura	Autoriza_UNIFESP.pdf	03/04/2019 23:26:42	Anna Alice Figueiredo de Almeida	Aceito
Declaração de Instituição e Infraestrutura	Autoriza_DepFonoUFPB.pdf	03/04/2019 23:25:29	Anna Alice Figueiredo de Almeida	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_Universal_Banco_de_vozes.pdf	01/03/2019 22:57:48	Anna Alice Figueiredo de Almeida	Aceito
Declaração do Patrocinador	Projeto_Universal2018_ContratoCNPq.pdf	01/03/2019 22:57:22	Anna Alice Figueiredo de Almeida	Aceito
Folha de Rosto	folha_de_rosto_CEP_assinada.pdf	01/03/2019 22:56:54	Anna Alice Figueiredo de Almeida	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

JOAO PESSOA, 06 de Maio de 2019

Assinado por:

Eliane Marques Duarte de Sousa
(Coordenador(a))



UNIVERSIDADE FEDERAL DA PARAÍBA
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
PROGRAMA ASSOCIADO DE PÓS-GRADUAÇÃO EM FONOAUDIOLOGIA

3. ARTIGO 2 - SUGERE SUBMISSÃO AO JOURNAL OF VOICE QUALIS A2 NA ÁREA 21 NO QUADRIÊNIO 2013-2016

Extração de medidas acústicas e perceptuais na variação das emoções

Héryka Maria Oliveira Lima ¹, Anna Alice Almeida²

¹ Fonoaudióloga, Mestranda do Programa Associado de Pós-Graduação em Fonoaudiologia, Departamento de Ciências da Saúde, Universidade Federal da Paraíba, João Pessoa, Brasil; ² Docente e pesquisadora do Departamento de Fonoaudiologia, Universidade Federal da Paraíba, João Pessoa, Brasil.

Resumo

Objetivo: Analisar parâmetros acústicos tradicionais do Banco de Vozes Brasileiro nas variações emocionais (EMOVOX-BR). **Desenho do estudo:** É um estudo transversal, descritivo e quantitativo. **Métodos:** Foram extraídos dados de 39 sinais de áudio, sendo 24 do sexo masculino e 15 sinais de áudio do sexo feminino, dentre as seis emoções básicas (alegria, medo, tristeza, raiva, surpresa) e a neutra. A extração ocorreu por meio de uso de script para extração de medidas de qualidade vocal. Foram extraídos parâmetros acústicos: frequência fundamental (F0) média, máxima, desvio padrão, mínimo em Hz, *jitter*, *shimmer*, intensidade; média e máximo de *extracion glotal noise* (GNE) em Hz. **Resultados:** A medida de qualidade vocal que melhor diferenciou as emoções foram as medidas de frequência fundamental: média, máxima, mínima e desvio padrão. Foi possível observar que as emoções se diferenciam entre si, e em relação às medidas acústicas, foi possível inferir correlações entre as medidas avaliadas. **Conclusão:** As medidas acústicas apresentam variações importantes na expressão das emoções nas vozes analisadas do EMOVOX-BR.

Palavras-chave: Fonoaudiologia, voz, emoções, comportamento, acústica.

Abstract

Objective: To analyze traditional acoustic parameters of the Brazilian Voice Bank in emotional variations (EMOVOX-BR). **Study design:** It is a cross-sectional, descriptive and quantitative study. **Methods:** Data were extracted from 39 audio signals, 15 female and 24 male, among the 6 basic emotions (joy, fear, sadness, anger, surprise) and neutral emission. The extraction took place through the use of a script to extract measures of vocal quality. Acoustic parameters were extracted: fundamental frequency (F0) average, maximum, deviation, minimum in Hz, jitter (%), shimmer (dB), intensity (dB); mean and maximum of glottal noise extraction (GNE) in Hz. **Results:** The vocal quality measure that best differentiated emotions were the measures of mean, maximum, minimum and deviation fundamental frequency. It was possible to observe that emotions differ from each other, and in relation to acoustic measures, it was possible to infer correlations between the measures evaluated. **Conclusion:** The acoustic parameters present important variations in the expression of emotions in the analyzed voices of the Brazilian Voice Bank with emotional variations.

Key Words: Speech therapy, voice, emotions, behavior, acoustics.

Introdução

As emoções são vivenciadas de diferentes formas por um indivíduo e está presente diariamente na vida dos seres humanos, elas podem ser expressas por diferentes formas, como por meio da face e da voz ¹. A voz desempenha uma importante função na expressão das emoções, sendo considerada o sistema de comunicação mais eficiente, rápido e natural do ser humano ^{2,3}. A interferência de aspectos emocionais na voz causa modificações anatomofisiológicas na produção vocal, essas modificações podem ser observadas por meios de análise acústica da voz. ^{3,4,5}

A análise acústica é um método objetivo de realizar avaliação da qualidade do sinal vocal, esse método fornece mensurações computadorizadas das propriedades acústicas da produção vocal de cada indivíduo. ^{6,9} É um método não invasivo, considerado objetivo por fornecer dados quantitativos e qualitativos. ^{7,8} A realização da análise acústica faz parte da avaliação multidimensional da voz, sendo responsável por gerar informações e traçados no formato de onda sonora. As medidas mais comuns na avaliação acústica da voz são relativas à extração de medidas como frequência fundamental, jitter, shimmer e medidas de ruído, a interpretação desses sinais permite entender a fisiologia da produção vocal. ^{6,9}

A análise do sinal de voz é um importante método para realizar o reconhecimento do sinal e interpretação das informações contidas nele ¹⁰. As medidas mais robustas para o reconhecimento das emoções por meio da voz são as medidas de *pitch*, formantes, medidas do trato vocal, coeficientes mel-cepstrais, intensidade do sinal e taxa elocução da fala. ^{13,14}

Buscas para identificar os sinais vocais nas variadas emoções, ocorrem comumente por meio de elaboração e análise de banco de vozes, existem aproximadamente 14 bancos de vozes publicados na literatura mundial, apenas seis são abertos a comunidade científica e dois são bancos de vozes infantis, no total, excluindo as dois bases infantis, destaca-se a existência de quatro banco de vozes aberto a comunidade científica¹⁹, eles são compostos por sinais vocais de indivíduos em idade adulta nas variações emocionais, nenhum dos 4 banco de vozes são brasileiros, todas foram desenvolvidos com intuito de analisar variações acústicas emocionais, dentre eles estão a base EMO-DB, foi desenvolvida na Universidade Técnica de Berlin, na Alemanha ¹⁵. A base IEMOCAP foi desenvolvida na Universidade do Sul da Califórnia, nos Estados Unidos¹⁶. A base SEMAINE é audiovisual e foi desenvolvida a partir de uma cooperação de Universidades da Inglaterra, da Holanda e da Alemanha ¹⁷. A base RECOLA foi desenvolvida na Suíça, e utilizou-se o idioma francês, apesar de terem participantes nativos de outros idiomas ¹⁸. Recentemente foi elaborado e validado o pioneiro

Banco de Vozes Brasileiro nas Variações das Emoções (EMOVOX-BR), que detém de 39 sinais vocais nas seis emoções básicas: alegria, medo, tristeza, raiva, surpresa, nojo e neutra.

19

Ainda não há um consenso a respeito de um atributo acústico que sinalize diferentes variações emocionais. Por tanto, o presente estudo poderá trazer contribuições acerca de um atributo que represente de forma significativa informações relacionadas às características acústicas próprias de cada emoção, extraída dos áudios que compõem o EMOVOX-BR. Isto posto, o objetivo deste estudo foi analisar parâmetros acústicos tradicionais do Banco de Vozes Brasileiro nas variações emocionais (EMOVOX-BR).

Metodologia

Esta é uma pesquisa transversal, descritiva e quantitativa, avaliada e aprovada pelo Comitê de Ética em Pesquisa do Centro de Ciências da Saúde de uma instituição de ensino superior no Nordeste do Brasil, sob número 3.304.419/2018. (ANEXO 1).

Amostra

Integraram esse estudo 39 sinais de áudio, sendo 24 do sexo masculino e 15 sinais de áudio do sexo feminino, dentre as seis emoções básicas, sendo composto por oito áudios da emoção surpresa, nas respectivas emoções tristeza, raiva e neutra sete áudios de cada emoção, cinco áudios da emoção medo, três áudios na emoção alegria e dois áudios na emoção nojo, os áudios foram extraídos do Banco de Vozes Brasileiro nas Variações das Emoções (EMOVOX-BR), desenvolvido por um laboratório da área de voz, do Nordeste do Brasil.

Esses 39 áudios apresentaram alto grau de concordância e valores satisfatórios na relação sinal-ruído (SNR). Os critérios de elegibilidade foram: obter valor acima de 30db no SNR, apresentar valor de Kappa acima de 0,7 de confiabilidade intra áudios, serem áudios coletados via smartphone e de forma line in, com o trecho de fala semi-espontânea “olha lá o avião azul” que compõe o *Consensus Auditory Perceptual Evaluation of Voice (CAPE-V)*²² validado recentemente para o português Brasileiro.²³

Materiais

Para obtenção das medidas acústicas dos 39 sinais analisados, foi utilizado script para extração de medidas de qualidade vocal, desenvolvido pelo Laboratório Integrado de Estudos da Voz (LIEV) da Universidade Federal da Paraíba, plugado ao software de acesso livre PRAAT (Paul Boersma e David Weenink, University of Amsterdam, Holanda), na versão 6.1.14. Foram extraídos os seguintes parâmetros acústicos de frequência fundamental (F0) média, máxima, desvio e mínimo em Hz, Jitter, shimmer, intensidade, média e máximo de *extracion glotal noise (GNE)*.

Procedimento de coleta de dados

As áudios analisados foram provenientes da base “EMOVOX-BR”, os 39 áudios foram normalizados no software Audacity, excluindo os três segundos iniciais e finais, visando evitar sinais irregulares, após esse processo, os áudios foram inseridos no aplicativo PRAAT,

inserido um áudio por vez, após esse processo, ainda no Praat foi adicionado o script para extração das medidas tradicionais de qualidade vocal, com esse processo, valores de saída foram extraídos e inseridos em uma planilha eletrônica no Microsoft Office Excel, para futura análise estatística dos dados.

Análise dos dados

Os dados foram digitados em planilha eletrônica no Microsoft Office Excel, para realizar a análise dos valores gerados pelo script PRAAT.

Foi realizado o teste de variância (ANOVA), com o objetivo de identificar quais as diferenças entre média e desvio padrão das medidas acústicas de cada emoção observada, após os resultados extraídos e identificados os grupos que se descartaram, foi realizado o teste post hoc de Turkey necessário para identificar qual o parâmetro acústico e a emoção se destacaram no grupo. O teste de correlação de Pearson foi utilizado para verificar e quantificar o grau de relacionamento linear entre as duas variáveis, observando as correlações entre as medidas acústicas e em que grau. Para a classificação dos coeficientes de correlação, nesta pesquisa foi adotado os valores de 0,1 a 0,3 representam uma correlação fraca; entre 0,4 e 0,6 indicam correlação moderada; e acima de 0,6 é possível afirmar que o grau de correlação entre as variáveis é forte.²⁴

Resultados

A Tabela 1 apresenta diferenças das medidas acústicas nas variadas emoções que foram testados usando medidas repetidas individuais do teste de variância (ANOVA) para cada parâmetro acústico. O resultado revela diferença significativa na comparação das médias das medidas de Frequência Fundamental (F0) nas variadas emoções, o valor de significância adotado $p < 0,05$.

Tabela 1- Comparação das medidas acústicas nas variadas emoções

Medidas Acústicas	Alegria		Medo		Tristeza		Raiva		Surpresa		Nojo		Neutra		P-Valor
	Média	DP	Média	DP	Média	DP	Média	DP	Média	DP	Média	DP	Média	DP	
F0 Max	129,6	37,3	252,03	45,2	104,9	24,1	134,98	47,9	141,1	69,5	110,7	28,2	99,3	20,7	0,001*
F0 Min	268,4	98,4	389,3	77,4	204,6	77,5	356,4	93,2	415,7	113,4	269,4	90,2	189,1	62,8	0,001*
F0 Med	234,0	86,0	311,7	59,1	153,4	43,3	222,4	55,9	248,6	88,8	153,3	7,67	147,5	42,1	0,001*
F0 Dp	22,4	10,2	27,1	15,2	18,5	12,2	46,3	16,6	50,2	16,4	27,5	4,3	20,0	13,9	0,001*
Jitter	8,37	13,6	10,06	9,64	8,28	11,27	11,29	11,52	3,6	6,6	13,7	3,2	6,2	10,7	0,768
Shimmer	60,6	62,3	69,0	41,4	56,7	72,4	75,3	76,0	25,1	46,2	92,7	32,9	48,8	85,9	0,775
Intensidade	550,2	83,8	510,5	153,8	503,8	123,7	581,9	134,4	595,3	76,7	453,5	90,1	587,9	109,7	0,492
GNE Med 1000Hz	0,276	0,028	0,271	0,49	0,289	0,29	0,274	0,25	0,278	0,021	0,293	0,005	0,274	0,027	0,917
GNE Max1000HZ	0,825	0,075	0,827	0,119	0,872	0,063	0,825	0,056	0,825	0,045	0,875	0,006	0,863	0,022	0,633
GNE Med 3000HZ	0,135	0,016	0,130	0,028	0,146	0,019	0,138	0,018	0,139	0,015	0,146	0,004	0,142	0,014	0,830

Legenda: F0= frequência fundamental; GNE= *glotal noise extracion*; max= máxima; Min= Mínimo; Med= médio, DP=Desvio Padrão. Teste ANOVA, P-Valor $< 0,05$.

A tabela 2 contempla os dados extraídos da análise do teste Post Hoc, realizado após a análise de variância (ANOVA). O teste post hoc apresentou-se necessário para identificar quais dos pares de grupos diferem, ou seja, qual o parâmetro acústico se destacou, bem como a emoção. Todas as emoções foram comparadas em pares, visando entender como elas se diferenciaram.

Tabela 2- Medidas de comparações múltiplas significantes

	Variáveis		Diferença média	Modelo Padrão	Significância
F0 Máximo	Medo	Alegria	122,4332733*	33,0434549	,013*
		Tristeza	147,1188114*	26,4937044	,001*
		Raiva	117,0522114*	26,4937044	,002*
		Surpresa	110,8889650*	25,7945387	,003*
		Nojo	141,2444400*	37,8560333	,012*
		Neutra	152,6982543*	26,4937044	,001*
F0 Mínimo	Neutra	Raiva	167,3167857*	47,5868262	,020*
		Medo	200,2478343*	52,1287563	,009*
		Surpresa	226,6575518*	46,0757463	,001*
	Tristeza	Raiva	151,8125286*	47,5868262	,045*
		Medo	184,7435771*	52,1287563	,019*
	F0 Médio	Medo	Tristeza	158,2860457*	36,5801807
Neutra			164,2529171*	36,5801807	,002*

F0 Dp	Tristeza	Raiva	-27,7266429*	7,7919992	,018*
		Surpresa	-31,6693214*	7,5445708	,003*
	Neutra	Raiva	26,2667571*	7,7919992	,029*
		Surpresa	30,2094357*	7,5445708	,006*

Legenda: Fo= Frequência fundamental, max= máxima; Min=Minimo;Med=médio, DP=Desvio Padrão. Teste post hoc de Tukey, diferença média é significativa no nível < 0,05.

É possível observar que a emoção medo foi a única que se diferenciou de todas as outras emoções na F0 máxima. Na F0 mínima a neutra se diferenciou das emoções medo, raiva e surpresa. Na F0 média a emoção tristeza se diferenciou das emoções medo e raiva. Quanto ao desvio padrão de F0, a emoção tristeza e a neutra se diferenciam da emoção raiva e surpresa.

A tabela 3 apresenta os achados do teste correlação de Pearson, que possui o objetivo de verificar possível correlação entre as variáveis.

Tabela 3- Correlações entre as Variáveis no Teste de Correlação de Pearson

VARIÁVEL	ESTATÍSTICA DO TESTE (valor da correlação de Pearson)	p-valor
F0 Máximo X F0 Mínimo	0,502	0,001*
F0 Máximo X F0 Médio	0,813	0,0001*
F0 Máximo X GNEMax1000Hz	-0,352	0,028*
F0 Máximo X GNEMed3000Hz	-0,318	0,049*
F0 Mínimo X F0 Médio	0,694	0,0001*
F0 Mínimo X F0 Dp	0,742	0,0001*
F0 Mínimo X GNEMax1000Hz	-0,459	0,003*
F0 Mínimo X GNEMed3000Hz	-0,377	0,018*
F0 Médio X F0 Dp	0,372	0,020*
F0 Médio X GNEMax1000Hz	-0,479	0,002*
F0 Médio X GNEMed3000Hz	-0,424	0,007*
Jitter X Shimmer	0,927	0,0001*
Jitter X Intensidade	-0,585	0,0001*
Shimmer X Intensidade	-0,687	0,0001*
Intensidade X GNEMed1000Hz	-0,570	0,0001*
Intensidade X GNEMax1000Hz	-0,325	0,043*
Intensidade X GNEMed3000Hz	-0,564	0,0001*

Legenda: Fo= Frequência fundamental, max= máxima; Min=Minimo;Med=médio, DP=Desvio Padrão. Teste de correlação de Pearson, P-Valor <0,05.

É possível observar que todas as variáveis demonstradas na Tabela três se correlacionam de forma significativa, e as variáveis com maior força de correlação foram: F0 máximo x F0 médio; F0 mínimo X F0 Dp; Jitter X shimmer; F0 Mínimo X F0 Médio; Shimmer X Intensidade. Todas essas correlações apresentaram força maior a 0,6 o que indica um grau forte de correlação entre as variáveis.

Discussão

A análise acústica da voz permite a identificação de traçados característicos na dinâmica de fala do indivíduo ²¹, durante a produção de fala é possível observar auditivamente mudanças no tom da voz, que ocorrem diversas vezes de acordo com o conteúdo de fala e a emoção expressa. ^{5,21,29,33}

Com avanço da ciência mudanças significativas na prática clínica fonoaudiológica ocorrem, como na avaliação da análise acústica da voz. Atualmente a extração de diversas medidas pode ser realizada por meio de um único script acoplado ao software de análise acústica. ²⁵ A contribuição dessas ferramentas mais objetivas na construção de conhecimento sobre variações emocionais comuns nos sinais vocais, podem favorecer a criação de padrões de reconhecimento das emoções. ^{4,9,14,20}

Em paridade com os estudos realizados anteriormente, o presente estudo buscou investigar parâmetros acústicos comuns nas variadas emoções no EMOVOX-BR, os dados relatados nesse estudo confirmaram que parâmetros acústicos são um fator importante para identificação das emoções por meio da análise acústica. ^{4,9,10,11,20,25}

Diante disso, buscou-se verificar diferenças entre as médias e desvio padrão das medidas acústicas de cada emoção, o parâmetro acústico que mais se diferenciou dos demais foram as medidas de frequência fundamental, essa por sua vez, está diretamente relacionado a função laríngea, portanto, alterações de entonação e fluxo de ar produzida pelo trato vocal devido ao estado emocional podem ser identificadas na análise acústica do sinal vocal. ^{26,27}

A F0 é considerada um parâmetro acústico promissor na identificação de falantes e geralmente não são afetados por conteúdo lexical de fala ^{21,29}, porém, alterações de estado emocional, ruído ambiental e estado de saúde podem afeta-lo. Portanto, a F0 se caracteriza como um parâmetro acústico capaz de diferenciar as emoções. ^{12,27,28}

Os achados acerca da frequência fundamental fortalecem a busca por identificar quais variações melhor se destacaram, as emoções medo, tristeza e a neutra foram as emoções que melhor se destacaram. Na frequência fundamental máxima a emoção medo foi a única que se destacou de todas as emoções alegria, tristeza, raiva, surpresa, nojo e neutra, esse achado está relacionado a variação da curva melódica e entonação de fala, sabe-se que modificações quanto a tensão das pregas vocais durante a vibração modificam a F0, quando mais alongadas as pregas vocais, mais rápido os ciclos glóticos ^{21,26,30}. Na emoção medo comumente observa-se pausas mais curtas, loudness forte e pitch agudo, esse por sua vez está diretamente ligado a F0, perceptualmente, esta impressão auditiva é chamada de pitch. ²¹

As medidas de jitter refletem o desvio global na emissão, já o shimmer estão relacionados a diminuição da resistência glótica, ambos apresentarão alta correlação entre si diretamente proporcional.^{25,30} O shimmer apresentou correlação inversamente proporcional com a variável intensidade, ou seja, quando menor a diminuição da resistência glótica maior a intensidade do sinal.

O jitter é considerado uma medida sensível às alterações de intensidade do sinal e presença de desvio da qualidade vocal, portanto, modificações como pausas curtas, alteração de intensidade e alterações na fluência, são marcadores comuns as emoções alegria, medo e raiva, esses traços podem ser identificados por meio da análise acústica.^{6,313,31, 32}

De forma geral, as medidas acústicas capazes de diferenciar as emoções são as medidas relacionada a fonte glótica, a F0 máximo diferenciou as 6 emoções básicas e a emoção medo foi a que mais se destacou, a F0 mínimo diferenciou a emoção tristeza, e a emissão neutra, a F0 média diferenciou as emoções medo e a F0 Dp diferenciou a emoção Tristeza e a emissão neutra. Por fim, é possível inferir que por meio de avaliação de sinais acústicos é possível identificar variações emocionais em falantes nativos do português brasileiro.

Conclusões

As variações emocionais apresentam repercussões nos parâmetros acústicos de qualidade vocal, a medida de Frequência fundamental se apresentou como um parâmetro importante para diferenciar as emoções. A emoção medo é a que melhor se diferenciou das demais emoções na F0 máxima. As características dos parâmetros acústicos apresentam variações importantes na expressão das emoções.

Referências

1. Belin, P., Fillion-Bilodeau, S., & Gosselin, F. The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, vol.40, p. 531-539, 2008.
2. EL AYADI, M.; KAMEL, M. S.; KARRAY, F. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, v. 44, n. 3, p. 572-587, 2011.
3. Florêncio, V. O.; Almeida, A. A.; Balata, P.; Nascimento, S.; Brockmann-bauser, M.; Lopes, L. W. Differences and Reliability of Linear and Nonlinear Acoustic Measures as a Function of Vocal Intensity in Individuals With Voice Disorders [published online ahead of print, 2021 Jun 8]. *J Voice*. 2021; S0892-1997(21)00144-2. Doi:10.1016/j.jvoice.2021.04.011
4. Almeida L N A, Lopes L W, Costa D B, Silva E G, Cunha G M S, Almeida A A F. Características vocais e emocionais de professores e não professores com baixa e alta ansiedade. *Audiol Commun Res*, 2014;19(2):179-85.
5. Braid, Antonio Cesar Morant. *Fonética Forense: tratado de perícias criminalísticas*. Campinas: Millenium, 2003.
6. Felipe ACNF, Grillo MHMM, Grechi TH. Normatização de medidas acausticas para vozes normais. *Rev. Bras. Otorrinolaringol*. V. 72, n5. Out 2006.
7. Araujo SA, Grellet M, Pereira JC. Normatização de medidas acústicas da voz normal. *Rev Bras Otorrinolaringol* 2002;68(4):540-4.
8. Finger LS; Cielo CA. Acoustic vocal modifications produced by reverse phonation. *Rev Soc Bras Fonoaudiol*. 2009;14(1):15-21.
9. Arevian AC; Bone D; Malandrakis N; Martinez VR; Wells KB; Miklowitz DJ; Narayanan S. Clinical state tracking in serious mental illness through computational analysis of speech. *PLOS ONE* | <https://doi.org/10.1371/journal.pone.0225695> January 15, 2020.
10. Low DM; Ghosh SS; Bentley KH. Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology*. 2020;5:96–116.
11. Fruhholz S; Grandjean D. Towards a fronto-temporal neural network for the decoding of angry vocal expressions. Elsevier Inc. All rights reserved. Doi:10.1016/j.neuroimage.2012.06.015.
12. Scherer KR. Vocal Communication of Emotion: A Review of Research Paradigms. *Speech Communication*. 2013;40(1):227–56.
13. Ververidis D, Kotropoulos C. Emotional speech recognition: Resources, features, and

methods. *Speech Communication*. 2006; 48:1162–81.

14. Patel S; Scherer KR; Bjorkner E; Sundberg J. Mapping emotions into acoustic space: The role of voice production. Elsevier B.V. All rights reserved. Doi:10.1016/j.biopsycho.2011.02.010.

15. Burkhardt F, Paeschke A, Rolfes M, Sendlmeier W, Weiss B. A Database of German Emotional Speech, in Proc. INTERSPEECH. Pp. 1517–1520, 2005.

16. Busso C, Bulut M, Lee CC, Kazemzadeh A, Mower E, Kim S, Chang JN, Lee S, Narayanan SS. IEMOCAP: Interactive Emotional Dyadic Motion Capture Database. *Language Resources and Evaluation*. V.42(4), p. 335–59, 2008.

17. Mckeown G, Valstar M, Cowie R, Pantic M, Schroder M. The SEMAINE 24 Database: Annotated Multimodal Records of Emotionally Colored Conversations between a Person and a Limited Agent. *IEEE Transactions on Affective Computing*. V.3(1), p. 5–17, 2012.

18. Ringeval F, Sonderegger A, Sauer J, Lalanne D. Introducing the recola multimodal corpus of remote collaborative and affective interactions. 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp. 1–8, 2013.

19. Lima, H M O; Almeida A A F. *Elaboração e Validação do EMOVOX-BR*. [Tese em Fonoaudiologia]. João Pessoa/PB: Universidade Federal da Paraíba, 2022.

20. Silva, W.J.N. *Marcadores vocais que discriminam pacientes com e sem depressão*. Tese-Universidade Federal da Paraíba. João Pessoa-PB.2019.

21. Kremer, Robinson Luis; GOMES, Maria Lúcia de Castro. A eficiência do disfarce em vozes femininas: uma análise da frequência fundamental. *Revel*, vol. 12, n. 23, 2014. [www.revel.inf.br].

22. ASHA – American Speech-Language-Hearing Association. Consensus auditory perceptual evaluation of voice (CAPE-V). *Rev. Soc. Bras. Fonoaudiol*, v. 9, n. 3, p. 187- 189, 2003.

23. Behlau, M.; Rocha, B.; Englert, M.; Madazio, G. Validation of the Brazilian Portuguese CAPE-V Instrument—Br CAPE-V for Auditory-Perceptual Analysis. *J Voice*, 2020 Aug 15; S0892- 1997(20)30257-5.

24. Dancy C, Reidy J. *Estatística sem matemática para psicologia: usando SPSS para Windows*. Porto Alegre: Artmed; 2006.

25. Lopes LW, Souza ESS, Silva ACF, Silva IM, Paiva MAA, Vieira VJD, Almeida AF. Medidas cepstrais na avaliação da intensidade do desvio vocal. *Codas*. 2018.

26. Lopes LW, Cavalcante DP, Costa PO. Intensidade do desvio vocal: integração de dados

perceptivo-auditivos e acústicos em pacientes disfônicos. *Codas* 2014;26(5):382-388.

27. Brockmann-Bauser M and Drinnan MJ. Routine acoustic voice analysis: time to think again? *Curr Opin Otolaryngol Head Neck Surg.* 2011;19(3):165- 70.

28. Godino-Llorente JI, Osma-Ruiz V, Sáenz-Lechón N, Gómez-Vilda P, Blanco-Velasco M, Cruz-Roldán F. Effectiveness of the glottal to noise excitation ratio for the screening of voice disorders. *J Voice.* 2010;24(1):47- 56. [Http://dx.doi.org/10.1016/j.jvoice.2008.04.006](http://dx.doi.org/10.1016/j.jvoice.2008.04.006). Pmid:19135854.

29. Künzle, Hermann J. Effects on voice disguise on speaking fundamental frequency. Birmingham: University of Birmingham Press, 2000.

30. Treole K, Trudeau MD. Changes in sustained production tasks among women with bilateral vocal nodules before and after voice therapy. *J Voice.* 1997;11(4):462-9. [Http://dx.doi.org/10.1016/S0892-1997\(97\)80043-4](http://dx.doi.org/10.1016/S0892-1997(97)80043-4). Pmid:9422281.

31. Behlau MS, Tosi O, Pontes P. Determinação da frequência fundamental e suas variações em altura (Jitter) e intensidade (Shimmer) para falantes do português brasileiro. *Acta A WHO*; 1985; 4:5-9.

32. Moreti F, Zambon F, Behlau M, Oliveira G. Equivalência cultural da versão Brasileira da Voice Symptom Scale – voiss. *J. Soc. Bras. Fonoaudiol.* 2011;23(4):398-400.

33. Ramos AKAR. Influência das emoções na voz de sujeitos investigados. Trabalho de conclusão de curso. Pontifca universidade católica de Goiais. Departamento de fonoaudiologia, 2021.

Anexo I

UFPB - CENTRO DE CIÊNCIAS
DA SAÚDE DA UNIVERSIDADE
FEDERAL DA PARAÍBA



Continuação do Parecer: 3.304.419

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1306092.pdf	03/04/2019 23:31:56		Aceito
Outros	Carta_resposta_CEP_UFPB.pdf	03/04/2019 23:31:25	Anna Alice Figueiredo de Almeida	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE2.pdf	03/04/2019 23:28:03	Anna Alice Figueiredo de Almeida	Aceito
Declaração de Pesquisadores	Atestado_MBehlau.pdf	03/04/2019 23:27:46	Anna Alice Figueiredo de Almeida	Aceito
Declaração de Instituição e Infraestrutura	Autoriza_UNIFESP.pdf	03/04/2019 23:26:42	Anna Alice Figueiredo de Almeida	Aceito
Declaração de Instituição e Infraestrutura	Autoriza_DepFonoUFPB.pdf	03/04/2019 23:25:29	Anna Alice Figueiredo de Almeida	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_Universal_Banco_de_vozes.pdf	01/03/2019 22:57:48	Anna Alice Figueiredo de Almeida	Aceito
Declaração do Patrocinador	Projeto_Universal2018_ContratoCNPq.pdf	01/03/2019 22:57:22	Anna Alice Figueiredo de Almeida	Aceito
Folha de Rosto	folha_de_rosto_CEP_assinada.pdf	01/03/2019 22:56:54	Anna Alice Figueiredo de Almeida	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

JOAO PESSOA, 06 de Maio de 2019

Assinado por:
Eliane Marques Duarte de Sousa
(Coordenador(a))



4. Considerações Finais

A elaboração e validação do banco pioneiro de vozes e variações emocionais Brasileiro- EMOVOX-BR, contribuiu para a ciência, uma vez que fornece sinais vocais de qualidade e confiabilidade, favoráveis para identificação das emoções.

Na etapa de validação, os 39 áudios hospedados no banco EMOVOX-BR possuem alta confiabilidade quanto à qualidade dos áudios, as juízas a partir da escuta dos áudios conseguiram identificar das emoções, valência e potência das emoções. As juízas fonoaudiólogas perceberam as modificações de voz e fala nas variadas emoções com boa concordância Interna. O parâmetro *pitch* se sobressaiu para identificação das emoções e emissão neutra, e *loudness* se apresentou como indicador da emoção raiva.

Quanto a identificação das medidas acústicas extraídas dos 39 áudios hospedado no banco de vozes brasileiro das variadas emoções, foi possível observar que os parâmetros acústicos se diferenciaram de acordo com as emoções avaliadas, a medida de Frequência fundamental se sobressaiu na identificação das emoções. A emoção medo é a que melhor se diferenciou das demais emoções na F0 máxima.

As identificações de características elencadas pelos juízes na análise perceptivoauditivo e as características dos parâmetros acústicos das vozes que compõem o banco de vozes brasileiro das variações emocionais apresentam variações importantes na expressão das emoções de falantes do português brasileiro.

REFERÊNCIAS

1. Adriano T, Arriaga P. Exaustão emocional e reconhecimento de emoções na face da voz em médicos. *SPPS*. 2016;17(1):97-104.
2. ALMEIDA, A. A. F.; BEHLAU, M.; LEITE, JR. Correlação entre ansiedade e performance comunicativa. *Rev Soc Bras Fonoaudiol*, v. 16, n.4, p.384-386, 2011
3. Almeida LNA, Lopes LW, Costa DB, Gonçalves Silva EG, Cunha GMS, Almeida AAF. Características vocais e emocionais de professores e não professores com baixa e alta ansiedade. *Scielo*, V.19(2), p.179-85, 2014.
4. Almeida AAF, Fernandes LR, Azevedo EHM, Pinheiro RSA, Lopes LW. Características vocais e de personalidade de pacientes com imobilidade de prega vocal. *CoDAS*. 2015; 27(2):178-185.
5. Barbosa RA, Friedman S. Emoção: efeitos sobre a voz e a fala na situação em público. *Distúrb Comum*, v.19(3), p. 325-336, 2007.
6. Behlau M. *Voz: O livro do especialista*. Vol. I. Rio de Janeiro: Editora Revinter, 2008.
7. Brasil CCP, Silva RM, Brilhante AVM, Melo AK, Batista MH. Entrelaçamento voz e emoção na percepção do docente sob a ótica da fenomenologia de Merleau-Ponty. *Rev. Interface*. Nov 2017. doi: 10.1590/1807-57622017.0344.
8. Burkhardt F, Paeschke A, Rolfes M, Sendlmeier W, Weiss B. A Database of German Emotional Speech, in *Proc. INTERSPEECH*. pp. 1517–1520, 2005.
9. Busso C, Bulut M, Lee CC, Kazemzadeh A, Mower E, Kim S, Chang JN, Lee S, Narayanan SS. IEMOCAP: Interactive Emotional Dyadic Motion Capture Database. *Language Resources and Evaluation*.V.42(4), p. 335–59, 2008.
10. Costa DB, Lopes LW, Silva EG, Cunha GMS, Almeida LNA, Almeida AAF. Fatores de risco e emocionais na voz de professores com e sem queixas vocais. *Rev. CEFAC*. 2013;15(4):1001-1010.
11. Ekman P. An argument for basic emotions. *Cognition and emotion*. 1992; 6, 169-200.
12. Felipe ACNF, Grillo MHMM, Grechi TH. Normatização de medidas acausticas para vozes normais. *Rev. Bras. Otorrinolaringol*. V. 72, n5. Out 2006.
13. Gomes VEFI, Costa DB, Lopes LW, Araujo RA, Almeida AAF. Symptoms and vocal risk factors in individuals with high and low anxiety. *Folia Phoniatica et Logopaedica*, 2018 [ahead of print].
14. Klostad CA, Anderson RC, Peters S. Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *P Roy Soc B-Biol Sci*. 2012; 279:2698–2704.
15. McKeown G, Valstar M, Cowie R, Pantic M, Schroder M. The SEMAINE Database: Annotated Multimodal Records of Emotionally Colored Conversations between a Person and a Limited Agent. *IEEE Transactions on Affective Computing*. V.3(1), p. 5–17, 2012.
16. Ringeval F, Sonderegger A, Sauer J, Lalanne D. Introducing the recola multimodal corpus of remote collaborative and affective interactions. *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp. 1–8, 2013.
17. SEIFERT, E.; KOLLBRUNNER, J. Stress and distress in non-organic voice disorder. *Swiss Med Wkly*, v. 135, n. 27-28, p. 387-97, 2005.
18. SUNDBERG, J. *A ciência da voz*. Editora da Universidade de São Paulo, 2015.

19. Trajano FMP, Almeida LNA, Araujo RA, Crisostomo FLS, Almeida AAF. Níveis de ansiedade e impactos na voz: uma revisão da literatura. *Distúrbios da Comunicação*. 2016; 28:423-33.
20. Tigue CC, Borak DJ, O'Connor JJM, Schandl C, Feinberg DR. Voice pitch influences voting behavior. *Evol Hum Behav*. 2012;33: 210–216.
21. VIEIRA, J. D. V. Análise de variações acústicas não estacionárias e seu efeito na detecção de múltiplas emoções e condições de estresse. Tese - Universidade Federal de Campina Grande. Campina Grande-PB. 2018.