



Reconhecimento de Emoções por Voz para Auxílio ao Diagnóstico de Transtornos Mentais

R Kingeski*, L A P Schueda, A S Paterno

Universidade do Estado de Santa Catarina, Joinville, Brasil

**rkingeski@hotmail.com*

Introdução. Emoções fazem parte do cotidiano de qualquer ser humano, todos os seres humanos possuem emoções positivas e negativas, que são reações naturais a estímulos, algumas emoções, consideradas negativas, tais como medo, raiva e tristeza, são mais comuns em pessoas com alguns transtornos mentais. Transtornos que possuem uma forte relação com emoções, por exemplo, são: o autismo, onde o indivíduo apresenta déficits na reciprocidade socioemocional, a esquizofrenia onde a pessoa tem sua expressão emocional reduzida, tanto facial quanto verbal, o transtorno de ansiedade generalizado, o estresse pós-traumático, que tem com um dos sintomas a incapacidade persistente de sentir emoções positivas (especialmente felicidade, alegria, satisfação ou emoções associadas a intimidade, ternura e sexualidade), depressão, transtorno de personalidade borderline, entre outros. Atualmente o Manual Diagnóstico e Estatístico de Transtornos Mentais DSM-V recomenda o uso de escalas transversais para fazer diagnósticos de transtornos mentais, essas escalas são basicamente questionários sobre sintomas associados aos transtornos. Um sistema de detecção de emoções em sinais de voz pode ser utilizado como complemento para analisar estados emocionais de pacientes com suspeita de doenças psiquiátricas, podendo também ser aplicado em sistemas com tecnologia vestível de monitoramento psicológico para pessoas que sofrem de doença mental, onde o reconhecimento de emoções pode fornecer aos terapeutas mais informações sobre as variações diárias do estado mental do paciente.

Materiais e Métodos. Para analisar um sinal de voz para identificar emoções existem diversas técnicas de processamento de sinal de voz. Esta análise é dividida em três partes: coleta de dados de voz, parâmetros de análise e classificação (DOI, 10.2478/v10177-012-0024-4). Este método é similar ao de identificação de voz. Neste trabalho foram coletados áudios de vídeos disponíveis na internet, de brasileiros nativos, em diversas situações que causaram a expressão de uma emoção. Para a coleta foram escolhidas as emoções básicas definidas por Paul Ekman, no estudo sobre emoções em expressões faciais, ele definiu as emoções básicas: neutra, medo, raiva, felicidade, tristeza, surpresa e nojo. Todos os algoritmos utilizados foram implementados no software MATLAB. E são baseados na revisão bibliográfica de algoritmos para a extração de parâmetros (DOI, 10.1109/MIPRO.2016.7522336). Para extração da frequência fundamental foi utilizado um janelamento de 60 milissegundos com sobreposição de 50 milissegundos. Foram extraídos os coeficientes LPC dos frames de voz, e então obtidos os valores do primeiro, segundo e terceiro formante. Também foi obtida a largura de banda de cada formante. E para energia de curto termo onde foi utilizado um janelamento de 20 milissegundos. No vetor de parâmetros, utilizado para classificação, foram utilizados: valor médio, valor mínimo, valor máximo e desvio padrão, para frequência fundamental, formantes e largura de banda de formantes, para a energia foi utilizado o valor médio quadrático. Totalizando 29 parâmetros. Foram coletados áudios de 5 pessoas, do sexo masculino, totalizando 206 áudios.

Resultados. Para a classificação foi utilizado o método de máquinas de vetores de suporte (SVM, do inglês *support vector machine*), com função de Kernel cúbica, para o treinamento da SVM foi



utilizado 155 áudios, de 4 locutores e para o teste foi utilizado 36 áudios de 1 locutor. A exatidão do treinamento foi de 84,5%, a matriz de confusão da Tabela 1 mostra o resultado.

Conclusões. Os resultados obtidos até o momento foram bons se comparadas ao estado da arte dos métodos de reconhecimento de emoções, e o número de áudios analisado, o próximo passo será aplicar estes métodos de extração de parâmetros e de classificação em um banco de dados que está em desenvolvimento, onde as emoções são induzidas em um estúdio de gravação. E fazer a classificação que não foi aqui utilizada, que é a classificação humana, onde pessoas escutam os áudios e informam qual a emoção ele consegue identificar em cada áudio. Por fim fazer uma comparação entre estes dois métodos de classificação em ambas as bases de dados obtidas. O projeto para a criação deste banco de dados por indução de emoções foi submetido ao comitê de ética com o título: Estudo de Reconhecimento de Emoções de um Banco de Dados de Voz em Idioma Português Brasileiro. CAAE: 83106617.5.0000.0118. Aprovado em 22/05/2018.

Tabela 1. Matriz de confusão da classificação dos parâmetros.

		Valor Previsto(%)						
		Medo	Neutro	Nojo	Raiva	Felicidade	Surpresa	Tristeza
Valor Verdadeiro(%)	Medo	81	9	0	9	0	0	0
	Neutro	45	55	0	0	0	0	0
	Nojo	0	0	97	3	0	0	0
	Raiva	8	0	4	69	12	0	8
	Felicidade	0	0	0	8	92	0	0
	Surpresa	0	0	25	0	0	75	0
	Tristeza	0	0	0	0	0	4	96

Palavras-chave. Reconhecimento de emoções, diagnóstico, transtorno mental, psiquiatria, processamento de sinais, voz.