



Identificação de Emoções por meio do Processamento Digital da Voz

T N Oriente^{1*}, I S Silva¹, Silvana L N C Costa¹

¹Instituto Federal da Paraíba, João Pessoa, Brasil

*taina.oriente@academico.ifpb.edu.br

Fundamentação, Motivação e Objetivo. As emoções exprimem a manifestação das reações dos seres humanos, muitas vezes influenciando na sua capacidade de socializar. O estudo das emoções, sobretudo baseado na voz, é pertinente para as interações homem-máquina (HMI), para a psiquiatria, ou até mesmo para a resolução de processos criminais e, nesse contexto, a identificação oriunda do julgamento humano é limitado e nutrido de subjetividade. Desse modo, este trabalho propõe o reconhecimento automático de emoções, baseado em parâmetros vocais, a partir de frases de locutores diversos, utilizando técnicas de redes neurais artificiais para classificação.

Métodos. Os discursos são oriundos do *Berlin Database Emotional Speech* (Emo-DB), o qual disponibiliza cerca de 800 arquivos de áudio no formato *.wav* divididos em 7 emoções distintas (felicidade, tristeza, tédio, desgosto, raiva, medo e neutro) e adquiridos a uma taxa de amostragem de 16 kHz. A Figura 1 ilustra o fluxograma do algoritmo desenvolvido. Após a aquisição do banco de dados, os sinais foram segmentados em trechos de 20 ms, realizando-se o cálculo dos parâmetros para cada trecho, a fim de aumentar a precisão do processo, visto que em um mesmo sinal há trechos com informação (discurso propriamente dito) e sem informação (pausas) que devem ser considerados. Os parâmetros calculados são: *Pitch* (0), que é o correlato perceptual da frequência fundamental que corresponde à frequência de vibração das pregas vocais; As quatro primeiras frequências formantes (1, 2, 3 e 4), que correspondem às frequências de ressonância do trato vocal; Taxa de cruzamento por zero (5); E energia (6), que é a intensidade do sinal de voz em decibéis (dB). A rede neural utilizada foi do tipo Perceptron de Múltiplas Camadas (MLP), cuja configuração é baseada na função de ativação logística e no algoritmo de retropropagação de erro, parametrizado com um erro mínimo de 10^{-4} e um número máximo de 500 iterações. Finalmente, a classificação foi realizada entre cada emoção e o estado neutro (6 classificações binárias), definindo 2/3 dos parâmetros para treino e 1/3 para teste, alimentando a rede com todas as combinações possíveis entre os 7 parâmetros (127 combinações).

Resultados Preliminares e Discussões. A Tabela 1 mostra os resultados das combinações de parâmetros que retornaram os melhores valores de Sensibilidade (*Sb*), Especificidade (*Sc*) e Acurácia (*Ac*) obtidos de cada classificação binária realizada. Os valores encontrados na coluna “Combinação” são os definidos para cada parâmetro na Seção de métodos. Nota-se que para as emoções mais distintas ao estado neutro, há um bom desempenho de classificação, enquanto que no caso contrário, este desempenho é reduzido significativamente.

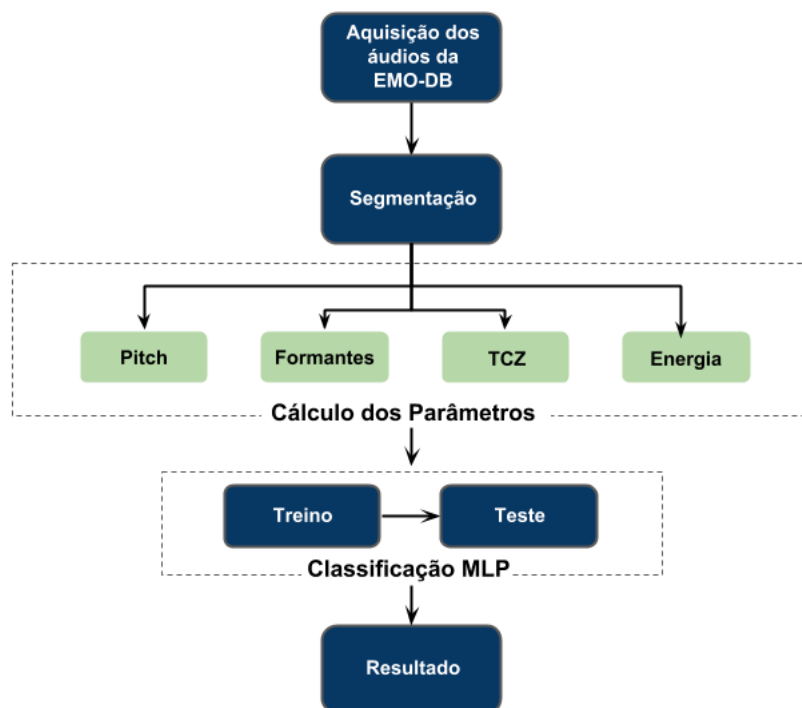
Tabela 1. Melhores resultados de cada classificação.

Classificação	Combinação	Sb (%)	Sc (%)	Ac (%)
Raiva vs Neutro	(2, 5)	92,31	100	95,45
Tédio vs Neutro	(0, 1, 2, 3, 6)	92,31	38,46	65,38
Desgosto vs Neutro	(0, 4)	100	100	100
Medo vs Neutro	(5)	95,65	69,56	94,44
Felicidade vs Neutro	(0, 2, 3, 4, 5)	91,30	78,26	85
Tristeza vs Neutro	(0, 3, 6)	85,71	85,71	87,5

No caso de Raiva vs Neutro, por exemplo, a rede conseguiu classificar corretamente em 95,45% dos testes, ao mesmo tempo em que essa classificação foi realizada corretamente em apenas 65,38% dos testes, para o caso de Tédio vs Neutro. Isso sugere que as emoções consideradas semelhantes, partindo de uma intuição meramente auditiva, apresentam parâmetros consideravelmente análogos, assim como o pensamento contrário também é indicado, sobretudo para aqueles parâmetros baseados em frequência, tais como o Pitch (0) e as Formantes (1, 2, 3 e 4).

Conclusões. Desse modo, é verificada a possibilidade do reconhecimento de emoções por meio dos parâmetros utilizados, visto que todos estes aparecem com uma frequência considerável nos melhores resultados, com destaque para o *Pitch* (0), que contribui na maioria das combinações, sugerindo a sua eficácia na classificação. É importante salientar que os resultados apresentados por este trabalho são preliminares, visto que o propósito inicial é evidenciar a possibilidade de identificação das emoções a partir dos parâmetros escolhidos.

Figura 1. Fluxograma do algoritmo desenvolvido.



Agradecimentos. Os autores agradecem ao IFPB pelo suporte financeiro.

Palavras-chave. Emoções; Voz; Parâmetros; Classificação; Rede Neural.