

## IDENTIFICAÇÃO DA MOTIVAÇÃO DO CHORO DE RECÉM-NASCIDOS UTILIZANDO ESPECTROGRAMAS

Gustavo Zanoni Felipe (PIBIC/CNPq), Yandre Maldonado e Gomes da Costa  
(Orientador), e-mail: [gzf1996@gmail.com](mailto:gzf1996@gmail.com), [yandre@din.uem.br](mailto:yandre@din.uem.br).

Universidade Estadual de Maringá / Centro de Tecnologia / Maringá, PR.

### Ciências Exatas e da Terra / Ciência da Computação

**Palavras-chave:** choro de recém-nascidos, espectrogramas, reconhecimento de padrões.

#### Resumo:

Perante adultos e crianças, entender a motivação de suas expressões diante de sensações como fome, dor e etc., não é uma tarefa difícil, já que as mesmas apresentam meios de comunicar tais sentimentos via sonoplastias comuns entre os seres humanos que são de fácil identificação. No caso dos recém-nascidos, que sem aprendizado suficiente para mostrar o que sentem, como um adulto ou uma criança mais velha faria, choram e fazem expressões faciais próprias, que podem variar de acordo com a idade e até mesmo com as condições de nascimento do bebê. Com isso, este trabalho objetivou identificar de forma automática se o choro de um recém-nascido era motivado por dor ou não. Para que isso fosse realizado, foram classificadas características extraídas a partir de espectrogramas gerados do sinal de áudio de gravações de bebês chorando. Ao fim deste trabalho, pode ser observado que os melhores resultados obtidos alcançaram 71,68% de acurácia, onde foi utilizado o descritor de textura LBP e a abordagem de extração global de características. Pode ser observado que o uso de técnicas de zoneamento e de combinação de características do meio acústico e visual não obtiveram um bom desempenho quando comparadas com as demais técnicas utilizadas.

#### Introdução

Para o ser humano, expressar sentimentos como por exemplo a dor é uma atividade comum, e pode ser facilmente identificada devido a associação feita à palavras e expressões verbais que simbolizam tal sentimento. Porém, quando o indivíduo ainda não sabe gesticular palavras concretas ou se expressar como uma criança ou um adulto faria, ele acaba procurando meios únicos de dizer às pessoas ao seu redor que algo não está certo.

Este é o caso dos recém-nascido, o choro é a primeira linguagem por meio da qual o recém-nascido se comunica e expressa suas necessidades. O choro é um sinal dinâmico e graduado, com características peculiares quando a dor é intensa. Logo, é relevante realizar a classificação utilizando de métodos de reconhecimento de padrões com o intuito de identificar se um bebê chora motivado por dor ou não. Já que, tais métodos não expõem fisicamente os recém-nascidos, além de se provarem eficientes e com resultados eficazes. Desta maneira, este trabalho objetivou identificar por

meio de características extraídas de um espectrograma gerado a partir da amostra de um áudio, se o choro de um recém-nascido é motivado pela ocorrência de dor ou não.

## **Materiais e métodos**

### *Base de Dados*

Para a realização deste projeto, foi utilizada uma base de dados à qual foi organizada em parceria com a pesquisadora da *State University of Missouri*, Professora Sheryl Brahnem. Esta é composta por 113 amostras de vídeo pré-segmentadas com 5 segundos de duração cada, onde é mostrada a expressão facial de recém-nascidos durante um momento de choro. A base é dividida em duas classes principais, uma onde os recém-nascidos têm o choro motivado pela ocorrência de dor (“com dor” ou *Pain*), enquanto na outra o choro não é motivado por dor alguma (“sem dor” ou *No Pain*). Do total de amostras, 42 pertencem à classe “com dor”, enquanto as demais 71 amostras pertencem à classe “sem dor”.

### *Metodologia*

A metodologia proposta neste trabalho, é baseada no trabalho de (COSTA, 2013), em que características extraídas de espectrogramas via descritores de textura eram utilizadas como entrada em um sistema de classificação. Realizando-se após uma fusão de classificadores para que a decisão final seja obtida. Para que os espectrogramas, representações gráficas geradas do conteúdo do sinal de uma amostra de áudio, fossem obtidas, foi utilizada a ferramenta *Sound EXchange* (SOX).

A extração de características foi realizada utilizando duas abordagens para zoneamento das imagens, sendo elas a extração global e a extração local. O método da extração global, baseia-se retirar as características do espectrograma como um todo, enquanto a extração local realiza a extração de características em faixas independentes da imagem, para que seja preservada a localidade da informação. Foram utilizados dois métodos de extração local: o Zoneamento Linear e o Zoneamento pela Escala Mel.

Para a extrair características, foram utilizados descritores de textura conhecidos na literatura para problemas de classificação de imagens, com o *Local Binary Pattern* (LBP), *Local Phase Quantisation* (LPQ) e o *Robust Local Binary Pattern* (RLBP). A classificação fora realizada utilizando a versão do algoritmo *Support Vector Machine* (SVM) implementada pela biblioteca LibSVM. Esta retorna uma estimativa de probabilidade de acerto para cada classe. Logo, diferentes classificadores, criados a partir de zonas diferentes da imagem, ou de descritores de textura diferentes, podem ser combinados com a finalidade de averiguar se a complementaridade das informações existentes produzem melhores resultados.

## **Resultados e Discussão**

Em um primeiro passo, foram extraídos os áudios de todas as amostras de vídeo, utilizando-se da ferramenta *ffmpeg*. Em seguida, foram gerados espectrogramas. Assim, procurou-se identificar inicialmente a quantidade de

*folders* que poderiam trazer melhores resultados nas taxas de classificação. Para isso, testes foram realizados variando-se tal quantidade entre três e seis *folders*. Em tais testes, utilizou-se do método de extração global e do descritor de textura LBP. Ressalta-se que o parâmetro referente ao limite inferior da amplitude do sinal representado no espectrograma também foi variado. Os melhores valores encontrados em tais testes podem ser observados na Tabela 1. Nestes, foi possível observar os limites inferiores da amplitude que se obtiveram melhores resultados (-90dB e -70dB), tanto quanto a quantidade de *folders* (4 e 5). Tais parâmetros foram então usados nos demais testes.

**Tabela 1** – Melhores resultados encontrados nos primeiros testes realizados.

Amplitude (dB)	Qtde. de Folds	Canal	Acurácia	F-Measure	Precision
<b>-90</b>	<b>4</b>	<b>Esquerdo</b>	<b>0,6903</b>	<b>0,6477</b>	<b>0,6812</b>
-70	4	Esquerdo	0,6549	0,6023	0,6215
-90	5	Direito	0,6549	0,6098	0,6211

Os testes iniciais apresentaram um problema, pois quase todas predições avaliadas faziam referência à classe “sem dor”. Sugerindo que o sistema não havia obtido aprendizado sobre a classe “com dor”. Assim, foram feitas alterações às amostras de áudio com o intuito de averiguar uma possível solução ao problema. Porém, testes realizados com tais variações apresentaram resultados insatisfatórios onde persistia-se o problema.

Em uma segunda tentativa, os tamanhos dos eixos do espectrograma foram aumentados para duas configurações: uma com o tamanho padrão do eixo das frequências e com o tamanho de 1500 pixels para o eixo do tempo (Config. #1) e, outra com eixos frequência e tempo com tamanhos 965 e 1333 pixels respectivamente (Config. #2). Tal tentativa alcançou resultados melhores que os obtidos até então, além de apresentar um avanço na solução do problema. Com isso, tal abordagem foi aplicada à novos testes e nestes, variou-se o descritor de textura. Os melhores resultados encontrados em tais testes, podem ser observados na Tabela 2. Os parâmetros utilizados para os resultados apresentados em tal tabela, foram utilizados nos demais testes.

**Tabela 2** – Melhores resultados encontrados nos testes com o valor referente aos tamanhos dos eixos alterados e utilizando de diferentes descritores de textura.

Amplitude (dB)	Qtde. de Folds	Configuração	Descritor	Canal	Acurácia	F-Measure	Precision
<b>-90</b>	<b>5</b>	<b>1</b>	<b>LBP</b>	<b>Direito</b>	<b>0,7168</b>	<b>0,6828</b>	<b>0,7040</b>
-70	4	2	RLBP	Esquer.	0,7008	0,6717	0,6938
-90	4	2	LBP	Direito	0,6903	0,6480	0,6761

Nos testes decorrentes, foram utilizadas técnicas de extração local e combinação de classificadores. Tais testes utilizaram do zoneamento linear (3, 4 e 5 zonas), zoneamento baseado na escala mel, combinação de classificadores entre as melhores predições obtidas, de ambos canais de áudio e diferentes descritores de textura. Porém, tais testes apresentaram acurácias inferiores à 68%, sendo que ocorria mais uma vez o problema já mencionado. Por fim, foram extraídas características “acústicas”, obtidas diretamente do sinal de áudio, utilizando as técnicas *Constant-Q Chromagram* (CQC) e o *Mel-Frequency Cepstral Coefficients* (MFCC). Tais características foram submetidas ao classificador, e os resultados produzidos foram combinados com os classificadores obtidos dos testes que envolviam características do meio visual (obtidos via descritores de textura em imagens de espectrograma). Nos resultados de tais testes, a maior acurácia encontrada, combinou características do descritor de textura LBP e do descritor acústico CQC, alcançando 65,49% utilizando da regra de fusão do máximo.

### Conclusões

O trabalho aqui desenvolvido objetivou desenvolver um sistema inteligente que identifica se o choro de um recém-nascido é motivado por dor ou não. Para isso foram utilizadas características extraídas de espectrogramas gerados a partir do sinal de uma amostra de áudio de um recém-nascido chorando. Ao final deste trabalho, diferentes testes foram realizados envolvendo a variação de diferentes parâmetros e abordagens. O melhor resultado encontrado a partir de tais testes alcançou acurácia de 71,68%, utilizando de características extraídas do meio visual via o descritor de textura LBP. Pode ser concluído também, que para o problema aqui abordado, técnicas de extrações locais de características e combinações de classificadores obtidos dos meios visual e acústico não apresentaram efetividade em seu uso. Visto que testes os quais as usaram, obtiveram resultados inferiores a outros encontrados.

### Agradecimentos

Ao professor Yandre Maldonado e Gomes da Costa pela orientação e pela oportunidade proporcionada. Ao CNPq pelo apoio financeiro.

### Referências

COSTA, Y. M. e G. **Reconhecimento de gêneros musicais utilizando espectrogramas com combinação de classificadores**. 2013. 125 f. Tese (Doutorado) – Programa de Pós-graduação em Informática, Universidade Federal do Paraná, Curitiba, 2013.