

CENTRO UNIVERSITÁRIO FEI
GUILHERME LUIZ PICOLOTTE DOS SANTOS
IZABELLA SIRQUEIRA
JÉSSICA DA SILVA COSTA
KIMBERLIN GONÇALVES CARDOSO
THALLES NUNES DA SILVA

**ANÁLISE E CLASSIFICAÇÃO DO CHORO DE RECÉM-NASCIDOS UTILIZANDO
MÉTODOS DE PROCESSAMENTO DE SINAIS**

São Bernardo do Campo

2021

GUILHERME LUIZ PICOLOTTE DOS SANTOS

IZABELLA SIRQUEIRA

JÉSSICA DA SILVA COSTA

KIMBERLIN GONÇALVES CARDOSO

THALLES NUNES DA SILVA

**ANÁLISE E CLASSIFICAÇÃO DO CHORO DE RECÉM-NASCIDOS UTILIZANDO
MÉTODOS DE PROCESSAMENTO DE SINAIS**

Monografia do Trabalho de Conclusão de Curso, apresentada ao Centro Universitário FEI como um dos requisitos necessários para a obtenção do título de Bacharel em Engenharia Elétrica. Orientado pelo Prof. Dr. Carlos Eduardo Thomaz e Co-orientado pelo Prof. Dr. Ivandro Sanches.

São Bernardo do Campo

2021

Ficha catalográfica

Folha de aprovação

A todos os nossos familiares e amigos que
acompanharam a trajetória desse sonho.

AGRADECIMENTOS

Agradecemos em primeiro lugar a Deus por ter-nos concedido sabedoria e força para que chegássemos até aqui. Dedicamos um agradecimento especial aos nossos orientadores, Prof. Dr. Carlos Eduardo Thomaz e Prof. Dr. Ivandro Sanches, por suas valiosas ideias, assistência, constante supervisão e apoio durante o desenvolvimento deste trabalho.

Também agradecemos a UNIFESP pela disponibilização da base de dados necessária para a realização de todo o estudo.

Por fim, expressamos a nossa gratidão e amor a todos os amigos e familiares que de alguma forma forneceram apoio e suporte durante os meses de desenvolvimento desta Monografia.

“ O maior inimigo do conhecimento não é a ignorância, mas a ilusão do conhecimento”

Daniel J. Boorstin

RESUMO

A análise do choro em crianças recém-nascidas é um estudo que revela uma preocupação associada à manifestação de dor, fome, sede ou algum incômodo em geral, pois diferentemente dos adultos, recém-nascidos são indivíduos que não conseguem de uma maneira verbal expressar suas vontades, sentimentos e dores, por isso, esses artifícios não-verbais como choro, inquietação e expressões faciais, podem ser usados para determinarmos suas necessidades. Ao longo dos anos, surgiram novas pesquisas e métodos com o intuito de avaliar tais comportamentos dos recém-nascidos, buscando identificar a forma de expressar pelo qual eles estão sentindo determinado desconforto e classificando-os com as possíveis causas destes. Segundo a literatura, os métodos abordados para realizar a análise do choro são fortemente dependentes de uma avaliação subjetiva realizada por humanos. Portanto, é necessário desenvolver métodos objetivos que permitirão a classificação automática do choro de dor. O estudo apresentado nesta monografia tem como objetivo a apresentação de um sistema de classificação e autenticação automática do choro de um recém-nascido, visando reconhecimento do recém-nascido que está emitindo o sinal do choro e diagnóstico para o choro de dor. Será realizado a validação do algoritmo simulado no software MATLAB, para aplicações de reconhecimento e autenticação do sinal de choro e a classificação do padrão do choro de dor e não dor.

Palavras-chave: Choro. Dor. Recém-nascido. Bebê. Acústica.

ABSTRACT

The analysis of crying in newborn children is a study that reveals a concern associated with the manifestation of pain, hunger or some discomfort in general, because unlike adults, newborn children are individuals who cannot verbally express their wills, feelings and pains, so these nonverbal devices such as crying, restlessness and facial expressions can be used to determine their needs. Over the years, new research and methods have emerged in order to evaluate such behaviors of babies, seeking to identify why they are feeling certain discomfort and classifying them with the possible causes of these. According to the literature, the methods approached to perform the crying analysis are strongly dependent on a subjective evaluation performed by humans. Therefore, it is necessary to develop objective methods that will allow the automatic classification of pain crying. The study presented in this monograph aims to present a classification system and automatic authentication of the cry of a newborn, aiming at the newborn recognition and diagnosis for the cry of pain. Validation of the simulated algorithm in MATLAB software will be performed, for applications of recognition and authentication of the cry signal and classification of the pattern of pain and non-pain cry.

Keywords: Cry. Pain. Newborn. Infant. Acoustic.

LISTA DE ILUSTRAÇÕES

Quadro 1 – Escala de dor infantil neonatal - NIPS	15
Quadro 2 – Indicadores comportamentais de dor infantil - BIIP	16
Quadro 3 – Escala de Sedação e Agitação da Dor Neonatal - N-PASS	17
Figura 1 – Diferenças anatômicas das vias aéreas superiores em um Recém-Nascido (RN)	18
Figura 2 – Características diferenciais da laringe de um RN	19
Figura 3 – Representação esquemática das fases do ciclo glotal com a ilustração dos movimentos das pregas vocais durante a formação dos pulsos glotais	20
Quadro 4 – Classificação da dor	31
Quadro 5 – Número de amostras de áudio por classe em base de dados	32
Figura 4 – Sinal no Tempo de uma amostra de dor aguda	33
Figura 5 – Sinal no tempo de uma amostra de dor não aguda	34
Figura 6 – Ilustração do processo <i>k-fold</i> para $k=5$	35
Figura 7 – Ilustração do processo <i>k-fold</i> para o agrupamento das amostras	40
Quadro 6 – Quantidade de acertos de acordo com número de gaussianas	40
Quadro 7 – Acurácia de acordo com número de gaussianas	41
Figura 8 – Acurácia parcial e total versus número de gaussianas adotadas	41
Figura 9 – Acurácia parcial (por classe) versus número de gaussianas adotados	42
Figura 10 – Sinal no Tempo da amostra Ruidosa	42
Figura 11 – Sinal no Tempo da amostra com possível fonação inspiratória	43

LISTA DE ABREVIATURAS

BIIP	Indicadores Comportamentais da Dor no Lactente (Behavioral Indicators of Infant Pain)
FDPS	funções densidade de probabilidade dos vetores acústicos (probability density function of acoustic vectors)
FFT	Transformada rápida de Fourier (Fast Fourier Transform)
GMM	Modelo de Mistura Gaussiana (Gaussian Mixture Model)
LPC	Codificação preditiva linear (Linear predictive coding)
MFCC	Coefficientes Mel-Cepstrais (Mel-frequency cepstral coefficients)
NIPS	Escala de Avaliação de Dor no Recém-Nascido (Neonatal Infant Pain Scale)
N-PASS	Escala Neonatal de Dor, Agitação e Sedação (Neonatal Pain Agitation and Sedation Scale)
RN	Recém-Nascido
SMSI	Síndrome da Morte Súbita Infantil

SUMÁRIO

1	INTRODUÇÃO	11
2	REVISÃO BIBLIOGRÁFICA	13
2.1	O QUE É DOR?	13
2.2	DOR EM RECÉM NASCIDOS	13
2.3	A FISIOLOGIA DO CHORO	18
2.4	PRINCIPAIS PADRÕES DE CHORO	21
2.5	TÉCNICAS DE ANÁLISE DE CHORO	21
2.6	ANÁLISE E AUTENTICAÇÃO DO SINAL DE CHORO	24
2.7	PROCESSAMENTO DE SINAIS	25
2.8	PRÉ-PROCESSAMENTO	25
2.9	IDENTIFICAÇÃO ATRAVÉS DO MODELO DE MISTURAS DE GAUSSI- ANAS	28
3	BASE DE DADOS	30
3.1	BASE DE DADOS UNIFESP	30
3.2	BASE DE DADOS iCOPE	31
4	FERRAMENTAS	33
4.1	WAVESURFER: FILTRAGENS INICIAIS	33
4.2	ALGORITMO 1: ANÁLISE ESPECTRAL E EXTRAÇÃO DAS CARACTE- RÍSTICAS	34
4.3	ALGORITMO 2: RECONHECENDO PADRÕES	34
4.4	VALIDAÇÃO CRUZADA	35
5	METODOLOGIA EXPERIMENTAL	37
5.1	FAMILIARIZAÇÃO E TESTES INICIAIS	37
5.2	IDENTIFICAÇÃO DO LOCUTOR DO CHORO	37
5.3	IDENTIFICAÇÃO DE PADRÕES DE DOR E NÃO DOR	39
6	COMPARAÇÃO DE METODOLOGIAS	44
6.1	ESTUDOS RELACIONADOS	44
6.2	TAXA DE ACERTO E EFICIÊNCIA	45
7	DISCUSSÃO E CONCLUSÃO FINAL	46
	REFERÊNCIAS	48

1 INTRODUÇÃO

O choro de um RN, intervalo de tempo que vai do nascimento até o momento em que a criança atinge 28 dias, pode ser definido como um fenômeno complexo que envolve uma combinação de vocalização, silêncio constritivo, tosse, engasgo e interrupções (ASTHANA; VARMA; MITTAL, 2015). Sendo esta sua primeira ferramenta de comunicação com o mundo externo e através dela sinalizam que estão com problemas ou necessidades, assim como os adultos realizam através da verbalização.

Pais e especialistas da área de cuidados de RNs aprendem a classificar os diferentes tipos de choro do bebê, baseando-se unicamente da percepção individual para a diferenciação auditiva e interpretação dos diversos tipos de choro (ASTHANA; VARMA; MITTAL, 2015).

Estudos anteriores apontam que o choro carrega informações essenciais que permitem a utilização deste parâmetro como um indicador de dor e também desenvolvimento neurológico em bebês recém-nascidos, e em um cenário clínico a rápida identificação de riscos, através de uma ferramenta de diagnóstico não invasiva, pode significar um tratamento mais assertivo, ou seja, bem-sucedido e eficaz na fonte da dor (GARCIA; REYES GARCIA, 2003). Os estudos também demonstraram que podemos distinguir diversos padrões e características singulares para as diferentes origens do choro, como dor, fome, medo etc. Além disso, através do choro podemos identificar vários estados patológicos, como lesão cerebral, fenda palatina, hidrocefalia e Síndrome da Morte Súbita Infantil (SMSI) (LEDERMAN, 2002).

O interesse crescente dos pesquisadores na relação presente entre as características do choro do bebê e as várias doenças ou necessidades apresentadas pelos RNs teve início na década de 1960. Os primeiros estudos envolveram o registro do choro infantil para análise em frequência, principalmente por meio de espectroscopia ou ultrassonografia. Muitos pesquisadores estabeleceram métodos de avaliação combinando certos padrões na expressão facial ou na frequência do choro com diferentes doenças e/ou necessidades (LEDERMAN, 2002). Nenhuma dessas avaliações foram aceitas como padrão, o que acarreta uma extrema necessidade de uma avaliação objetiva e de métodos de classificação automática do choro do bebê (RUSU et al., 2015).

Diante do exposto acima, o trabalho desenvolvido nesta monografia tem como objetivo a apresentação de um sistema de classificação e autenticação automática do choro de um RN, visando reconhecimento do locutor e diagnóstico para o choro de dor. O algoritmo será simu-

lado no software MATLAB, utilizando uma linguagem de alto nível. Outros softwares também integram a análise, como o editor de áudio Wavesurfer.

2 REVISÃO BIBLIOGRÁFICA

Esta seção visa apresentar e discutir as principais definições e técnicas empregadas na análise do choro de dor e na autenticação de um sinal de voz.

2.1 O QUE É DOR?

A palavra dor tem origem no latim *dolor*, sofrimento. Do inglês, o termo *pain*, deriva das palavras *poine* do grego e *poena* do latim, que significam pena, punição. Atualmente, a dor é compreendida como um importante componente de defesa corporal diante de um ambiente hostil, pois constitui um sinal de alarme que visa proteger o organismo de possíveis danos (LENT, 2008).

A dor é definida como uma experiência sensitiva e emocional desagradável associada, ou semelhante àquela associada a uma lesão tecidual real ou potencial, sendo a informação sensorial captada pelo sistema nervoso periférico e transmitida ao sistema nervoso central (SRINIVASA et al., 2020). Tal definição reconhece o fenômeno da dor como uma experiência visto que envolve tanto as dimensões sensorial-discriminativa como afetivo-motivacional. O componente sensorial-discriminativo se refere à percepção e detecção do estímulo nocivo quanto a sua intensidade, localização, duração, portanto, relacionado com dor aguda. Por outro lado, o componente afetivo-motivacional diz respeito à relação entre as reações emocionais desencadeadas pelo processo doloroso, portanto, relacionado com dor crônica (LENT, 2008).

A ausência de comunicação verbal não nega a possibilidade de um indivíduo estar vivenciando dor e, muito menos, a sua necessidade de um tratamento adequado. Os RNs são mais suscetíveis aos eventos de interpretação incorreta sobre os eventos dolorosos e demais necessidades, devido a sua inabilidade de comunicação verbal. (DA SILVA, 2017).

2.2 DOR EM RECÉM NASCIDOS

Os RNs eram tidos como insensíveis a dor até a década de 1970 e, por conta desta concepção, vários procedimentos médicos dolorosos como inserção de cateter ou até mesmo cirurgias, eram realizadas sem o cuidado de usar algum tipo de anestesia ou analgesia adequada (BRANCO; FEKETE; SOUZA RUGOLO, 2006).

Devido ao crescente enfoque da literatura realizado nas últimas duas décadas sobre a percepção de dor no período neonatal, é possível estudar a reação a dor apresentada pelo recém-nascido mediante a um estímulo doloroso.

Como um reflexo a dor, o RN apresenta rigidez do tórax, flexão e extensão das extremidades, movimentos específicos das mãos e alterações na mímica facial. Após a punção capilar no calcanhar, sabe-se que em 0,3 segundos o paciente retira a perna não puncionada, em 0,4 segundos retira a perna puncionada e, em 1,8 segundos, chora (GUINSBURG; CUENCA, 2010). Inicialmente a resposta a dor do RN é um reflexo involuntário e conforme o sistema nervoso amadurece e à medida que as respostas aos estímulos dolorosos vão evoluindo, o choro de dor deixa de ser somente um ato de reflexo e torna-se mais organizada e consistente (BRANCO et al., 2006).

A partir das respostas faciais, corporais e do choro, é possível decodificar essa linguagem da dor. Com essa visão, foram desenvolvidas escalas unidimensionais, que avaliam a resposta comportamental à dor, e ferramentas multidimensionais, que incluem uma combinação de parâmetros objetivos e subjetivos relacionados à resposta à dor exibida pelo recém-nascido.

Dentre todas as escalas existentes, algumas se apresentam com maior frequência na literatura. Elas são a Escala de Avaliação de Dor no Recém-Nascido (NIPS, Neonatal Infant Pain Scale), a Indicadores Comportamentais da Dor no Lactente (BIIP, Behavioral Indicators of Infant Pain) e a Escala Neonatal de Dor, Agitação e Sedação (N-PASS, Neonatal Pain Agitation and Sedation Scale), descritas a seguir e apresentadas nos quadros 1, 2 e 3, respectivamente.

- a) NIPS (Neonatal Infant Pain Scale): A Escala de Avaliação de Dor no Recém-Nascido é composta por cinco parâmetros comportamentais e um indicador fisiológico, avaliados antes, durante e após procedimentos invasivos agudos em recém-nascidos a termo e pré-termo. A maior dificuldade reside na avaliação do parâmetro “choro” em pacientes entubados – nessa situação, dobra-se a pontuação da mímica facial, sem avaliar o “choro”. A interpretação dessa escala se dá somando os valores de pontuação de acordo com os aspectos apresentados no quadro. Define-se dor quando a pontuação é ≥ 4 , considerando um somatório das pontuações da tabela. (GUINSBURG; CUENCA, 2010).
- b) BIIP (Behavioral Indicators of Infant Pain): A escala Indicadores Comportamentais da Dor no Lactente é uma modificação recente do Sistema de Codificação Facial do Recém-Nascido (NFCS) que inclui o estado de alerta do recém-nascido e a movimentação das mãos, tornando a avaliação comportamental mais especí-

fica e inserida na interação entre paciente e ambiente. A interpretação dessa escala se da somando os valores de pontuação de acordo com os aspectos apresentados no quadro. Defini-se dor quando a pontuação é ≥ 5 (GUINSBURG; CUENCA, 2010).

- c) N-PASS (Neonatal Pain Agitation and Sedation Scale): Escala Neonatal de Dor, Agitação e Sedação: Trata-se de uma escala válida e confiável, que possui variáveis fisiológicas e comportamentais, desenvolvida para avaliar dor aguda e prolongada (crônica e/ou contínua) e sedação. É composta por duas medidas de escore: dor/agitação e sedação e, em cada uma, cinco critérios são avaliados: choro/irritabilidade, estado comportamental, expressão facial, tônus das extremidades e sinais vitais. A interpretação dessa escala se da somando os valores de pontuação de acordo com os aspectos apresentados no quadro. Define-se dor quando a pontuação é ≥ 3 (BALDA; GUINSBURG, 2018).

Quadro 1 – Escala de dor infantil neonatal - NIPS

NIPS	0	1	2
Expressão Facial	Relaxada	Contraída	-
Choro	Ausente	"Resmungos"	Vigoroso
Respiração	Relaxada	Diferente do basal	-
Braços	Relaxados	Flexão ou extensão	-
Pernas	Relaxadas	Flexão ou extensão	-
Estados de alerta	Dormindo ou calmo	Desconfortável	-

Fonte: os Autores "adaptado de" Guinsburg e Cuenca, 2010

Quadro 2 – Indicadores comportamentais de dor infantil - BIIP

Estado de sono/vigília		
BIIP	Pontos	Definição
Sono Profundo	0	Olhos fechados, respiração regular, ausência de movimentos das extremidades.
Sono Ativo	0	Olhos fechados, contração muscular ou espasmos/abalos, movimento rápido dos olhos, respiração irregular.
Sonolento	0	Olhos fechados ou abertos (porém com olhar vago, sem foco), respiração irregular e alguns movimentos corporais.
Acordado/Quieto	0	Olhos abertos e focados, movimentos corporais raros ou ausentes.
Acordado/Ativo	1	Olhos abertos, movimentos ativos das extremidades.
Agitado/Chorando	2	Agitado, inquieto, alerta, chorando.
Face e Mãos		
BIIP	Pontos	Definição
Fronte Saliente	1	Abaulamento e presença de sulcos acima e entre as sombrancelhas.
Olhos espremidos	1	Compressão total ou parcial da fenda palpebral.
Sulco nasolabial aprofundado	1	Aprofundamento do sulco que se inicia em volta das narinas e se dirige à boca.
Estiramento horizontal da boca	1	Abertura horizontal da boca acompanhada de estiramento das comissuras labiais.
Língua tensa	1	Língua esticada e com bordas tensas.
Mão espalmada	1	Abertura das mãos com os dedos estendidos e separados.
Mão fechada	1	Dedos fletidos e fechados fortemente sobre a palma das mãos formando um punho cerrado/ mão fechada.

Fonte: os Autores “adaptado de” Guinsburg e Cuenca, 2010

Quadro 3 – Escala de Sedação e Agitação da Dor Neonatal - N-PASS

	Sedação		Sedação/Dor	Dor/Agitação	
	-2	-1	0/0	1	2
Choro / Irritabilidade	Não chora c/ estímulo doloroso.	Resmunga/ chora com estímulo doloroso.	Sem sinais de sedação ou dor.	Irritado ou episódios de choro consolável.	Choro agudo silencioso contínuo. Não é consolável.
Padrão Comportamental	Não acorda com estímulo, sem movimento espontâneo.	Acorda breve com estímulo, raro movimento espontâneo.	Sem sinais de sedação ou dor.	Inquieto, se contorce. Acorda com frequência.	Arqueia o corpo, fica chutando. Acordado constante ou não acorda nem se move (não está sedado).
Expressão Facial	Boca caída e aberta, sem mímica.	Mínima expressão facial com estímulo.	Sem sinais de sedação ou dor.	Qualquer expressão de dor intermitente.	Qualquer expressão de dor contínua.
Tônus de extremidade	Sem reflexo de preensão, flácido.	Reflexo de preensão fraco, tônus muscular ↓	Sem sinais de sedação ou dor.	Mãos serradas ou espalmadas de modo intermitente. Tônus corporal relaxado.	Mãos cerradas ou espalmadas de forma contínua. Tônus corporal tenso.
Sinais Vitais: FC, FR e SatO₂	Sem Δ após estímulo, hipoventilação ou apneias.	Δ<10% com estímulo.	Sem sinais de sedação ou dor.	↑ 10-20% em relação ao basal SatO ₂ 76-85% com estímulo, rápida recuperação.	↑ 20% em relação ao basal SatO ₂ <75% com o estímulo, lenta recuperação, sem sincronia com o ventilador.

Fonte: os Autores “adaptado de” Balda e Guinsburg, 2018

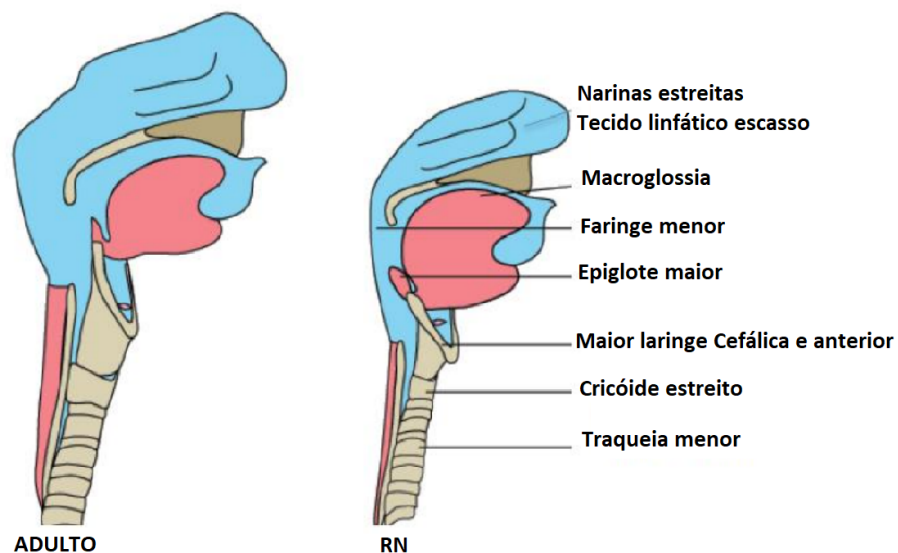
2.3 A FISILOGIA DO CHORO

Para compreender os significados das várias características do sinal do choro e para ser capaz de analisar e diferenciar os padrões deste, é essencial familiarizar-se com a fisiologia relacionada ao mecanismo de produção do choro.

O choro no RN pode ser definido como um fenômeno complexo que ocorre durante a expiração da respiração, quando todos os músculos supralaríngeos, laríngeos, respiratórios, músculos faciais, da boca e do tronco devem estar funcionando corretamente (BRANCO et al., 2006). Toda essa ação do sistema fisiológico é decorrente do sistema nervoso, que assegura ao recém-nascido toda a estabilidade e coordenação laríngea necessária para a operação.

No RN, as estruturas do trato vocal apresentam características particulares devido à configuração do trato vocal nesta faixa etária. Neste momento, a laringe está posicionada no ponto superior do pescoço e as pregas vocais possuem em torno de 2,5 a 3 milímetros (mm) de comprimento. A mucosa é espessa, o ligamento vocal não está completamente desenvolvido e a laringe apresenta o formato de funil com cartilagens delicadas e flexíveis (BRANCO et al., 2006). Toda essa estrutura pode ser observada através da figura 1.

Figura 1 – Diferenças anatômicas das vias aéreas superiores em um RN

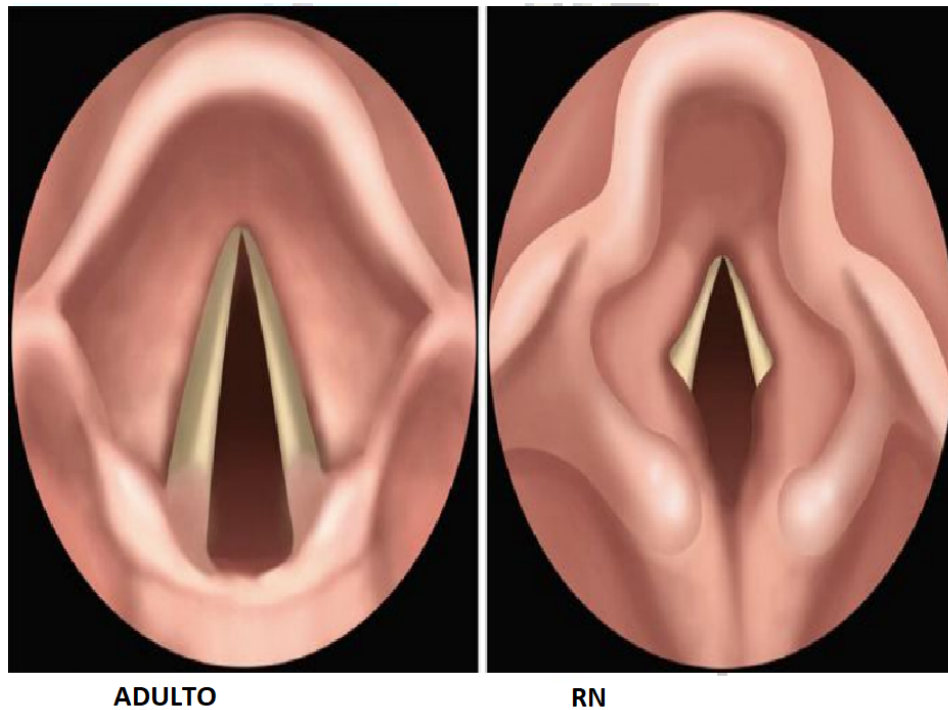


Fonte: os Autores “adaptado de” (REVISTA... , s.d.)

O mecanismo de produção do choro em RNs se assemelha com o processo de produção da fala em adultos. A primeira fase do mecanismo de produção do choro é iniciada no cérebro através de estímulos externos ou internos (fome, dor, desconforto, etc.). Na segunda fase o

cérebro produz uma série de comandos através do sistema nervoso que resultam na produção acústica de sons no nível fisiológico. Este processo continua através da ejeção de ar dos pulmões para o trato vocal produzindo fluxo e pressão de ar capazes de excitar o mecanismo vibratório das pregas vocais. (SARASWATHY et al., 2012).

Figura 2 – Características diferenciais da laringe de um RN

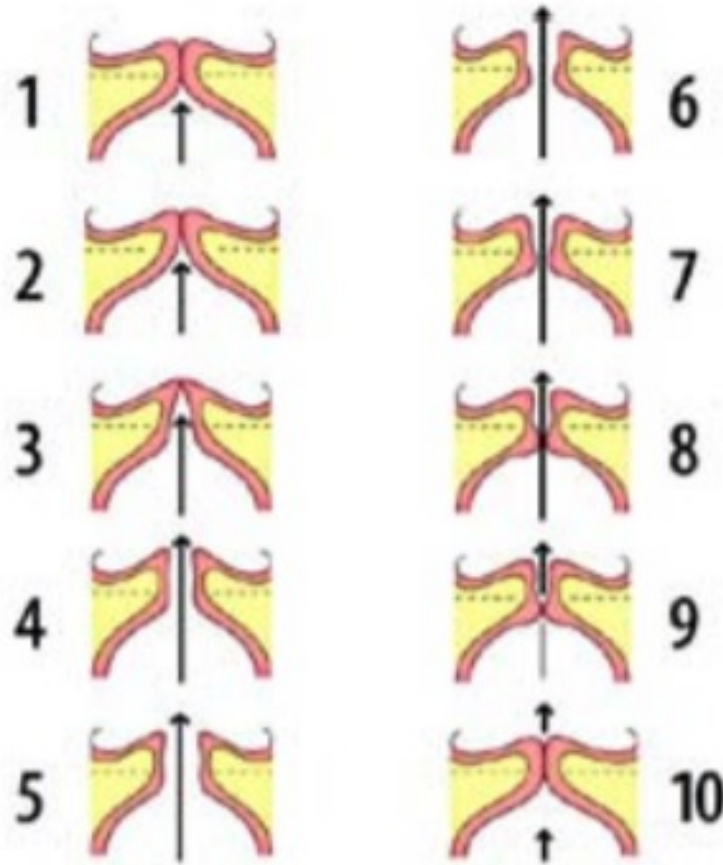


Fonte: os Autores “adaptado de” (REVISTA..., s.d.)

Durante a fonação, ao sair dos pulmões, a corrente de ar encontra as pregas vocais aduzidas, gerando um conflito entre o fluxo aéreo e as forças da musculatura laríngea. As forças que mantêm as pregas vocais fechadas são vencidas, momentaneamente, pela pressão subglótica, fazendo com que o ar passe pela região glótica. A passagem do ar gera uma pressão negativa perpendicular ao fluxo aéreo, o efeito de Bernoulli, que, associado à força elástica dos tecidos, traz a mucosa das pregas vocais de volta à linha média. Com o novo aumento da pressão subglótica, inicia-se o novo ciclo vibratório. Assim sendo, cada ciclo glótico é formado por quatro etapas: fase fechada, fase de abertura, fase aberta e fase de fechamento. Estas etapas se repetem consecutivamente, transformando o fluxo de ar, originalmente contínuo, em pulsos denominados pulsos glotais, gerando um som sensível à audição humana (BAECK, 2006). Os cortes da laringe podem ser vistos na figura 3.

Estudos anteriores apontam que o controle do choro é adquirido após um mês do nascimento. Durante o período de controle involuntário o choro do recém-nascido é um reflexo

Figura 3 – Representação esquemática das fases do ciclo glotal com a ilustração dos movimentos das pregas vocais durante a formação dos pulsos glotais



Fonte: os Autores “adaptado de” (VANZELLA, 2006)

de vários estados psicofisiológicos como fome, desconforto e dor. Conforme o sistema nervoso é desenvolvido o choro torna-se uma forma real para expressar sentimentos e desconfortos (BRANCO et al., 2006).

Os primeiros estudos que descreveram os padrões acústicos dos diferentes tipos de choro do recém-nascido surgiram na década de 60 e através da extração dos valores da frequência fundamental tornou-se possível a obtenção de informações referentes ao funcionamento das estruturas vocais do recém-nascido frente a estímulos externos (BRANCO et al., 2006).

A frequência fundamental é caracterizada pela tensão exercida nas pregas vocais e pelos músculos da laringe. O som laríngeo produzido nas pregas será modificado pelas estruturas do trato vocal e através dos valores da frequência fundamental é possível mapear os diferentes tipos de padrão do choro e as suas variações situacionais. Este mapeamento é essencial para

o auxílio da avaliação da dor e no tratamento que será abordado para avaliar a dor do recém-nascido (BRANCO et al., 2006).

2.4 PRINCIPAIS PADRÕES DE CHORO

Muitos pesquisadores investigaram as expressões de dor no recém-nascido, sejam essas faciais ou acústicas. Wasz-Hockert, Michelsson e Lind (1985) realizaram uma análise acústica espectrográfica em 148 amostras de sinais de choro e identificaram 2 padrões básicos de choro (fome e dor):

- a) O choro de dor em um recém-nascido saudável é caracterizado pela frequência máxima fundamental de 650 Hz e frequência mínima de 400 Hz. A melodia do sinal do choro de dor é ascendente e descendente, apresenta queda na frequência de sustentação e cada segmento possui duração média de 2.5 segundos. A maioria dos choros de dor é tensa, e segundo o pesquisador tal característica pode ser explicada pelo efeito do estresse provocado pelo estímulo doloroso, que aumenta a tensão dos músculos do aparelho vocal, ou seja, dos músculos laríngeos e das pregas vocais, elevando os formantes das frequências e alterando as suas intensidades. Portanto, o nível de estresse na situação de dor pode alterar o padrão do choro e, em consequência, influenciar a resposta do cuidador em relação à criança. A tensão e a estridência características da vocalização frente à experiência dolorosa são essenciais para o julgamento do adulto, pois chamam atenção e ajudam a discriminar uma situação de dor muito intensa ou pouco intensa. Nos recém-nascidos prematuros, o choro de resposta ao estímulo de dor também apresenta frequência elevada, bifonação e sobreposição dos primeiros picos espectrais (BRANCO; FEKETE; SOUZA RUGOLO, 2006).
- b) Nas amostras de sinais de choro de fome foi identificada uma média para a frequência fundamental máxima de 550 Hz e o mínimo de 390 Hz.

2.5 TÉCNICAS DE ANÁLISE DE CHORO

Esta sub-seção visa apresentar e discutir os principais tipos de análises aplicadas no estudo do sinal do choro.

a) ANÁLISE AUDITIVA

A escuta diagnóstica, embora imprecisa, é o meio mais rápido e imediato para a análise do choro. Wasz-Hockert (WASZ-HOCKERT; MICHELSSON; LIND, 1985) mostrou que os dois padrões básicos de choro podiam ser identificados auditivamente. O estudo desenvolvido por este pesquisador relatou que o choro foi reconhecido melhor por adultos que tiveram experiências anteriores com choro infantil, como é o caso da equipe de enfermarias pediátricas e parteiras. Também foi constatado que o treinamento aumentou a capacidade de reconhecer diferentes tipos de choro: após um treinamento de aproximadamente 2 horas, 82 pediatras puderam diagnosticar choro normal versus choro patológico com muita precisão. Também foi possível diagnosticar através do choro alguns tipos de patologia específica, embora com menos precisão.

Além disso, verificou-se que as mães podiam reconhecer o choro do bebê após ouvir apenas um choro. O estudo incluiu choro de bebês com asfixia, lesão cerebral, hiperbilirrubinemia e síndrome de Down (WASZ-HOCKERT; MICHELSSON; LIND, 1985).

Embora tenha sido constatado que o treinamento melhorou a capacidade de reconhecer os principais padrões de choro, a análise auditiva ainda fornece apenas uma fração das informações contidas no sinal de choro, informações que podem ser obtidas através de técnicas de diagnóstico mais sofisticadas.

b) ANÁLISE NO DOMÍNIO DO TEMPO

No domínio do tempo, o parâmetro de duração consiste na forma mais intuitiva de descrever o choro. Entre as mais relevantes formas de descrever esse parâmetro, podemos destacar: a duração total do episódio de choro, a duração de trechos específicos, duração das pausas e a latência entre o estímulo e o início do choro, além de também ser possível observar a energia do choro (LEDERMAN, 2002).

Esta técnica tem a vantagem de ser relativamente fácil de operar, e é barata, confiável e fácil de inspecionar visualmente. No entanto, a quantidade de informações são limitadas usando este método.

c) ANÁLISE NO DOMÍNIO DA FREQUÊNCIA

A análise do domínio de frequência permite obter uma representação grosseira das características do espectro de frequência de um som. Os dispositivos

empregados nesta análise, através de frequências de entradas selecionadas, medem a magnitude média e fornecem uma exibição visual da magnitude relativa. Assim, é possível comparar essa magnitude relativa com diferentes frequências (LEDERMAN, 2002).

Como implícito no próprio nome, esses dispositivos apenas dão informações sobre a magnitude relativa de várias faixas de frequência.

d) ANÁLISE ESPECTROGRÁFICA

A análise espectrográfica integra a análise no domínio do tempo e da frequência para produzir uma análise espectral que mostra a distribuição de energia tanto no domínio da frequência quanto no domínio do tempo. Este tipo de análise tem servido como uma técnica muito importante e útil em diversas áreas de processamento de sinal (WASZ-HOCKERT; MICHELSSON; LIND, 1985). A maioria dos estudos sobre o choro do bebê realizados ao longo dos anos utilizaram como metodologia a análise espectrográfica do sinal de choro. Embora seja uma ferramenta útil, esta técnica ainda requer uma análise manual e cuidadosa, sua dependência por um especialista na análise final acaba por tornar este método inadequado para um banco de dados em grande escala e para o uso clínico (WASZ-HOCKERT; MICHELSSON; LIND, 1985).

e) ANÁLISE COMPUTACIONAL

Todos os métodos de análise discutidos anteriormente possuem limitações significativas. Em todos, o processo de extração das características do sinal de choro é um processo difícil e complexo. O desenvolvimento de métodos de análise baseados em computadores e softwares nos últimos anos tornou possível desenvolver sistemas automáticos que podem extrair informações acústicas relevantes de forma rápida e confiável (LEDERMAN, 2002). A análise computacional permite a determinação mais precisa de informações acústicas e extração de informações que, de outra forma, seriam impossíveis de obter. Para o desenvolvimento desta pesquisa será utilizado um algoritmo de autenticação de voz para realizar a verificação e identificação do recém nascido que está emitindo o sinal do choro analisado, e também, identificar se é um choro de dor ou não dor.

2.6 ANÁLISE E AUTENTICAÇÃO DO SINAL DE CHORO

Reconhecer a fala de um indivíduo é um importante traço no que diz respeito a comunicação de humano para humano. Na relação entre bebês e adultos, essa tentativa de verbalização acontece de maneira intencional através do choro da criança, e os adultos de forma instintiva tendem a reconhecer o apelo no choro de seus bebês (SANTOS, 2000). Baseado na importância desse reconhecimento, ao longo do tempo diversos aplicativos foram desenvolvidos com o intuito de verificar, autenticar e identificar essa tentativa de comunicação entre indivíduos. A abrangência desse reconhecimento automático vai desde sua utilização no universo dos eletrônicos portáteis como os assistentes pessoais baseados em reconhecimento de fala (HANSEN; HASAN, 2015) à utilização em aplicativos como babás eletrônicas que identificam e monitoram o sono dos bebês e, caso algum ruído sonoro associado ao choro do bebê seja detectado, utiliza-se características para auxiliar na validação de sinais de voz, identificando o choro da criança e imediatamente um alerta é enviado para os pais.

Os primeiros registros sobre o estudo de reconhecimento de voz surgiram por volta dos anos 1950, onde a análise era feita por uma pessoa especialista que examinava a amostra do sinal de voz e, por meio de comparações com outros sinais, decidia para qual indivíduo o sinal seria associado. A autenticação de voz é uma ferramenta que vem evoluindo de um processo manual para um processo automatizado que envolve algoritmos de alta complexidade (MELO, 2014), que combinada com a experiência de um profissional consequentemente permite uma análise mais assertiva se comparada à análise puramente humana. A utilização de um discurso falado ou a tentativa de vocalização de um bebê através do choro, vem sendo utilizado como um parâmetro de identificação de um indivíduo e cada vez mais viável devido ao progresso tecnológico e avanços extremamente consideráveis no quesito de processamento de sinais (MELO, 2014), o que permite a análise, síntese e modelagem dos sinais que posteriormente poderão ser usados em alguma aplicação específica, como o tratamento e autenticação de amostras de choro de recém-nascidos na identificação dos bebês combinada com uma análise mais aprofundada destinada a classificação do grau de dor que está ocasionando o choro e com o auxílio de um profissional treinado diagnosticar o grau de dor que a criança está submetida.

A extração dos parâmetros da fala é uma etapa fundamental para os sistemas de reconhecimento de voz, pois através dela consegue-se identificar a energia do sinal, sua frequência fundamental e sua forma estrutural (VIANA, 2013). Há muitas características que compõe a vocalização de um bebê e que o torna reconhecível, principalmente por se tratar de sua relação

familiar com seus pais. Considerando esse tipo de reconhecimento não há a necessidade de um treinamento consciente, pois se trata de algo intrínseco da convivência e da comunicação. No reconhecimento automático a análise da fala com auxílio de um especialista é complementada pela tomada de decisão com métodos computacionais (HANSEN; HASAN, 2015).

As variações oriundas do orador estão intrinsecamente relacionadas à maneira como o discurso é entonado, incluindo seu esforço vocal, linguagem, dialetos, emoções etc. Já na conversação (observada no momento da aquisição das amostras de áudio dos bebês), a variabilidade se torna uma mescla de vários fatores, pois, o diálogo ou uma comunicação em grupo além de conter as variações intrínsecas de cada indivíduo, também contém ruídos com a mistura das entonações. Além de fatores diretamente relacionados a criança, um fator extremamente importante na aquisição das amostras de áudio é o meio tecnológico pelo qual a fala/comunicação está acontecendo seja por meio de microfones, *headsets*, gravações em aparelhos eletrônicos e também como estão sendo reproduzidas (HANSEN; HASAN, 2015).

2.7 PROCESSAMENTO DE SINAIS

A autenticação do sinal do choro de dor é realizada através do processamento do sinal, e suas principais etapas são: pré-processamento, identificação, comunicação e aplicação (RODRIGUES, 2017). Esta pesquisa focará exclusivamente nas duas primeiras etapas, que serão discutidas a seguir.

2.8 PRÉ-PROCESSAMENTO

A etapa de pré-processamento consiste basicamente em: Conversão A/D, filtragem e extração dos parâmetros acústicos. Esta etapa pode ser ainda subdivida em três micro etapas: a pré-ênfase, a análise espectral e a extração das características. (RODRIGUES, 2017).

a) Pré-ênfase

Primeiramente é realizada a conversão de um sinal analógico para digital. De forma simples e objetiva esse processo pode ser dividido em três etapas: Amostragem, quantização e codificação.

– Amostragem

Selecionar pontos (Amostras) desse sinal e armazená-los, tornando o tempo discreto. A frequência de amostragem mínima deverá ser duas vezes a frequência máxima do sinal analógico.

- Quantização

Tornar a amplitude discreta, através da divisão em níveis, visivelmente o sinal ficará quadrado.

- Codificação

Cada nível identificado no item anterior será codificado binariamente, sendo importante a escolha adequada da quantidade de bits necessários para a codificação.

Após a conversão, é realizada a filtragem do sinal, com um filtro de primeira ordem simples, este processo reduz os efeitos de ruídos aumentando a magnitude de alguns componentes de frequência, equalizando o espectro da mensagem, onde todas as componentes demandem a mesma largura de banda.

b) Análise Espectral

Após a filtragem, é realizada a análise espectral, que consiste em converter o sinal, que está no domínio do tempo, para o domínio da frequência. A conversão para o domínio da frequência é realizada através da divisão do sinal em quadros de duração fixa, este processo é denominado janelamento.

Para realizar a análise espectral dois métodos são muito utilizados (VIANA, 2013), eles são:

- Métodos de bancos de filtros de Transformada rápida de Fourier (FFT, Fast Fourier Transform)
- Codificação preditiva linear (LPC, Linear predictive coding)

c) Extração das características

Por fim, concluindo a etapa de pré-processamento, têm-se a extração das características do sinal, sendo a obtenção dos Coeficientes Mel-Cepstrais (MFCC, Mel-frequency cepstral coefficients) um dos métodos mais utilizados.

- Coeficientes Mel-Cepstrais

A escala Mel foi desenvolvida experimentalmente na década de 1940, seu objetivo era identificar como diferentes frequências eram interpretadas

pelo aparelho auditivo humano e descrever uma relação entre a frequência real e o que era interpretado (MELO, 2014).

Matematicamente a relação pode ser descrita na equação 1:

$$F_{mel} = \frac{1000}{\log(2)} \log \left[1 + \frac{F_{Hz}}{1000} \right] \quad (1)$$

onde:

- i. F_{mel} representa a frequência resultante em escala Mel;
- ii. F_{Hz} representa frequência medida em Hz.

Segundo Rodrigues (2017) esta relação é aproximadamente linear de 0 a 1000 Hz e para frequências superiores a 1000 Hz, a relação pode ser descrita de forma logarítmica. Na prática, a conversão do sinal para a escala Mel é implementada utilizando filtros passa-banda, de resposta triangular com espaçamento e largura determinados por um intervalo constante de frequência Mel (RODRIGUES, 2017).

Após a conversão, calcula-se o cepstrum, sendo este o resultado do cálculo da transformada inversa de Fourier do logaritmo do espectro de sinal estimado. Em linhas gerais o cepstrum consiste na representação do espectro de potência de curto prazo de um som, com base em uma transformada de cosseno linear de um espectro de potência, uma escala logarítmica é usada para posicionar as bandas de frequência (RODRIGUES, 2017).

A equação 2 descreve, matematicamente, o processo de obtenção dos coeficientes Mel-Ceptrais.

$$c_n = \sum_{k=1}^K \log(S_k) \cos \left[n \left(k - \frac{1}{2} \frac{\pi}{K} \right) \right], n = 1, \dots, L \quad (2)$$

onde:

- i. L representa o número de coeficientes;
- ii. S_k representa os coeficientes de potência da saída do k-ésimo filtro.

Com a determinação dos coeficientes, é possível iniciar a macro etapa de identificação, descrita na próxima seção.

2.9 IDENTIFICAÇÃO ATRAVÉS DO MODELO DE MISTURAS DE GAUSSIANAS

A etapa de identificação consiste em realizar o reconhecimento da informação baseada em representações existentes dos padrões observados. Há duas vertentes importantes, matematicamente falando, por trás do processo de identificação, sendo elas: a extração dos vetores característicos, com as informações dos parâmetros intrínsecos da fala de cada locutor, formados pelos coeficientes Mel-Cepstrais, que contribui para formar um modelo semelhante ao aparelho responsável pela fonação humana e a correlação desses vetores de coeficientes em uma combinação linear de funções de densidade, que dá origem ao Modelo de Mistura Gaussiana (GMM, Gaussian Mixture Model), que utiliza a probabilidade na identificação dos modelos correlacionados com o conjunto de vetores e, por meio da iteração do algoritmo se o locutor tiver o modelo mais próximo será então identificado (CARDOSO, 2009).

GMMs têm sido uma potente ferramenta, utilizada para extração de recursos de dados de fala.

O modelo em questão é um modelo estocástico, ou seja, que modela classes.

Para que seja realizada a classificação, faz-se necessário o cálculo da chamada probabilidade a posteriori, que pode ser calculada através da fórmula de Bayes (HANSEN; HASAN, 2015), como mostrado na equação 3:

$$P(C_i|o) = \frac{p(o|C_i)P(C_i)}{p(o)} \quad (3)$$

Onde:

- a) C_i representa uma classe (Para um modelo com um total de M classes, $i = 1...M$);
- b) o representa uma observação, ou seja, vetor acústico observado;
- c) $p(o|C_i)$ é a chamada probabilidade condicional a priori;
- d) $P(C_i)$ é a chamada probabilidade a priori da classe i .

O elemento observado corresponde a uma sequência de vetores acústicos, logo se considerarmos uma sequência de T vetores, com a ocorrência de cada observação como um evento independente, pode-se construir uma regra de decisão para o problema através da maximização da probabilidade a posteriori (HANSEN; HASAN, 2015), matematicamente, essa regra pode ser representada na equação 4:

$$a_k = \arg \max_{1 \leq i \leq M} \prod_{t=1}^T p(o_t|C_i)P(C_i) \quad (4)$$

Onde:

- a) a_k representa a regra de decisão ;
- b) C_i representa uma classe, para um modelo com um total de M classes, $i = 1 \dots M$;
- c) o_t representa uma sequência de vetores acústicos observados;
- d) $p(o_t|C_i)$ é a chamada probabilidade condicional a priori;
- e) $P(C_i)$ é a chamada probabilidade a priori da classe i.

A decisão é realizada, utilizando as funções densidade de probabilidade dos vetores acústicos, também conhecidas como funções densidade de probabilidade dos vetores acústicos (FDPS, probability density function of acoustic vectors). Com elas é possível realizar a modelagem do sistema de reconhecimento (HANSEN; HASAN, 2015).

A estimativa priori das funções é realizada através de um algoritmo de re-estimação de parâmetros, sendo o algoritmo de Baum-Welch, o mais usual, também conhecido como *forward-backward* ou algoritmo de avanço-retorno. Através desse algoritmo, é possível então, estimar os parâmetros de médias, variâncias e pesos das misturas de Gaussianas para cada função de densidade de probabilidade (HANSEN; HASAN, 2015).

Logo, matematicamente, os GMM podem ser definidos segundo a equação 5:

$$p(o_t|C_j) = p(o_t|\lambda_j) = \sum_{i=1}^I \Pi_i N(o_t; \mu_i, \Sigma_i) \quad (5)$$

A probabilidade condicional $p(o_t|C_j)$ é substituída por $p(o_t|\lambda_j)$.

onde:

- a) λ_j é um modelo GMM para a classe C_j ;
- b) Π_i representa o peso de cada gaussiana na mistura;
- c) o_t representa uma sequência de vetores acústicos observados;
- d) $N(o_t; \mu_i; \Sigma_i)$ representa uma gaussiana multivariada, com vetores de médias e variâncias μ_i e Σ_i , respectivamente I número de gaussianas.

O modelo fornecerá, então, a chamada probabilidade condicional, que é calculada na equação 5 e que será usada no processo de decisão apresentado na equação 4. (HANSEN; HASAN, 2015)

Em suma, o GMM se baseia na suposição de que cada elemento do conjunto se origina a partir de um componente da mistura com uma determinada probabilidade. Assim, ao inferir os parâmetros da mistura, essa probabilidade pode ser usada para associar cada elemento ao componente com maior probabilidade de o ter gerado (HANSEN; HASAN, 2015).

3 BASE DE DADOS

Os próximos tópicos irão apresentar os bancos de dados utilizados durante o desenvolvimento deste trabalho.

3.1 BASE DE DADOS UNIFESP

Para a aplicação dos algoritmos em MATLAB no reconhecimento de locutor, foi utilizada uma base de dados composta por 19 recém-nascidos que participaram de um projeto em desenvolvimento entre a FEI e a UNIFESP o FAPESP 2018/13076-9.

As amostras foram extraídas durante a realização da punção nos pacientes. Todos os bebês que participaram deste estudo foram declarados como saudáveis, de ambos os sexos e variadas origens étnicas, estando dentro dos limites aceitáveis de peso e altura. Infelizmente, não foi possível a obtenção de mais dados que pudessem indicar e classificar nominalmente estes bebês.

Ao final de cada teste, o especialista presente durante a execução do exame classificou de acordo com a sua experiência e conhecimento a intensidade da expressão de dor naquele momento, vários especialistas foram designados para a avaliação. Cada teste realizado era classificado por um especialista diferente com experiência adequada para o reconhecimento da intensidade de dor no RN.

O quadro 4 apresenta as principais informações referentes aos recém-nascidos que compõem a base de dados.

Quadro 4 – Classificação da dor

Identificação Padrão	Intensidade da expressão de dor
Bebê 1	7
Bebê 2	6
Bebê 3	8
Bebê 4	7
Bebê 5	8
Bebê 6	sem identificação
Bebê 7	8
Bebê 8	10
Bebê 9	6
Bebê 10	4
Bebê 11	4
Bebê 12	8
Bebê 13	8
Bebê 14	8
Bebê 15	7
Bebê 16	3
Bebê 17	3
Bebê 18	6
Bebê 19	7

Fonte: os Autores

3.2 BASE DE DADOS iCOPE

Para realizar a validação do algoritmo na identificação do padrão de choro de dor e não dor, foi utilizada a iCOPE Database. A base de dados iCOPE surgiu inicialmente pelo estudo da diferença de desempenho da classificação SVM de dor neonatal através das expressões faciais dos RNs (BRAHNAM et al., 2006). O banco original contém, ao total, 204 fotos gráficas de 25 RNs, com idade de 18 a 36 horas. As amostras em imagem foram obtidas através dos registros em vídeo realizados nos RNs durante a punção no calcanhar e outras experiências de desconforto (mudança de berço, troca de fralda ou fome) (BRAHNAM et al., 2006).

A partir das amostras em vídeo foram geradas 113 amostras de áudio com duração de 10 segundos. Esse conjunto de dados também são disponibilizados a outros pesquisadores através da iCOPE Database. Dentre as 113 amostras em áudio, 42 foram classificadas como dor, e 71 como sendo de não dor (FELIPE et al., 2019). A relação final entre a classificação das amostras podem ser verificadas no quadro 5.

Os áudios agrupados na classe de não dor, são referente aos registros em vídeos dos momentos de desconforto experimentados pelos RNs. Já os áudios agrupados na classe do dor, são referentes aos registros em vídeos do momento em que foi realizado a punção no calcanhar do RN.

Quadro 5 – Número de amostras de áudio por classe em base de dados

Classe	Quantidade
DOR	42
N-DOR	71
TOTAL	113

Fonte: os Autores

Segundo o estudo de Classificação SVM (BRAHNAM et al., 2006), o processo de coleta de dados obedeceu aos protocolos e diretrizes éticas de pesquisa envolvendo seres humanos da universidade de Missouri e do Hospital St. John's Health System. Apenas mães que tiveram partos sem complicações foram abordadas para participarem do estudo. Todos os RNs participantes do estudo eram Caucasianos, igualmente divididos entre os sexos (13 meninos e 12 meninas) e sem complicações de saúde. Para maiores informações referente a base, o leitor pode consultar o estudo na íntegra.

É válido ressaltar que devido ao estudo que originou a base utilizada neste trabalho ser focado na análise da expressão facial de dor em RNs, algumas amostras de áudios foram geradas sem possuir sinal de choro. Isso deve-se ao fato de que nem todos os RNs expressam dor somente através do choro, a expressão facial também é um meio de comunicação importante para a análise da presença de dor.

Outro ponto digno de nota é que as 113 amostras de áudios não significam 113 RNs. Após a extração das amostras de áudio dos vídeos também foi realizado o seccionamento dos áudios em pequenos trechos de 5 segundos.

Para validação do método aplicado e para a comparação com os resultados obtidos em outros estudos relacionados ao tema, todas as amostras de áudio classificadas como dor e como não dor foram mantidas.

4 FERRAMENTAS

Nos próximos tópicos serão apresentadas as principais ferramentas e os algoritmos utilizados para a filtragem, recombinação, extração e análise dos parâmetros.

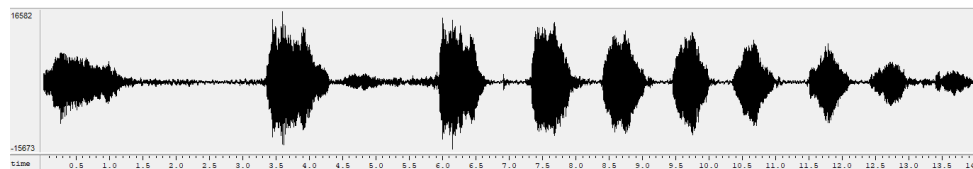
4.1 WAVESURFER: FILTRAGENS INICIAIS

A primeira ferramenta para o desenvolvimento deste trabalho, como um todo, experimentalmente falando, foi o Wavesurfer. Este programa foi inicialmente empregado na transformação dos formatos de áudios, de MP3 para WAV. Posteriormente foram realizadas análises no domínio do tempo, onde foram identificadas amostras que estavam fora do padrão pré-estabelecido, ou seja, haviam amostras sem som, outras com um intervalo de tempo menor do que o mínimo aceitável e por fim outras com ruídos de agentes externos, como mães falando ao fundo, barulhos de equipamentos, etc. De certa, as amostras passaram pelo processo de pré-enfase manualmente, onde através da filtragem da base melhorou-se a relação de Sinal Ruído, maximizando o desempenho do sistema, garantindo que através da eliminação de discontinuidades e quaisquer efeitos, seja obtido um conjunto de dados contendo somente informações relevantes ao processo de identificação.

Exemplificando as análises no domínio do tempo através do programa Wavesurfer têm-se:

A Figura 4 retrata a resposta no domínio do tempo de um bebê que apresenta uma expressão grau 8 de dor, de acordo com o critério de classificação do especialista na base disponibilizada pela UNIFESP e utilizada na identificação de locutor.

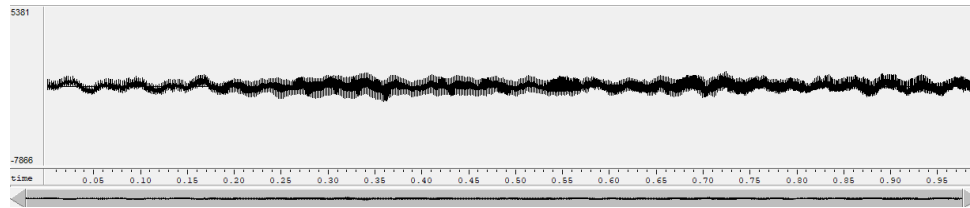
Figura 4 – Sinal no Tempo de uma amostra de dor aguda



Fonte: os Autores

A Figura 5 retrata a resposta no domínio do tempo de um bebê que expressou grau 3 de dor, de acordo com o critério de classificação do especialista na base disponibilizada pela UNIFESP e utilizada na identificação de locutor.

Figura 5 – Sinal no tempo de uma amostra de dor não aguda



Fonte: os Autores

4.2 ALGORITMO 1: ANÁLISE ESPECTRAL E EXTRAÇÃO DAS CARACTERÍSTICAS

Após as conversões e filtragens, foram iniciadas duas importantes etapas, a análise espectral e a extração dos parâmetros acústicos. Teve-se o emprego de um novo algoritmo, que é executável em duas ferramentas, sendo elas: O MATLAB e o OCTAVE. Este algoritmo utilizou um dos métodos mais populares para a extração das características, em reconhecimento de voz, a obtenção dos MFCC. É importante mencionar que a seleção de um método eficiente para a extração dos parâmetros acústicos do sinal é uma das etapas mais importantes do pré-processamento dos dados (RODRIGUES, 2017).

Ao executar o algoritmo, primeiramente teve-se a verificação da duração mínima, áudios com menos de 1 segundo não possuíam dados suficientes para uma análise assertiva e foram descartados, após a verificação foi realizada a conversão do áudio, do formato WAV para HTK, onde o sinal foi convertido do domínio do tempo, para o domínio da frequência. Nesta etapa foram realizadas as extrações das características do sinal através da obtenção dos coeficientes mel-cepstrais correspondentes.

4.3 ALGORITMO 2: RECONHECENDO PADRÕES

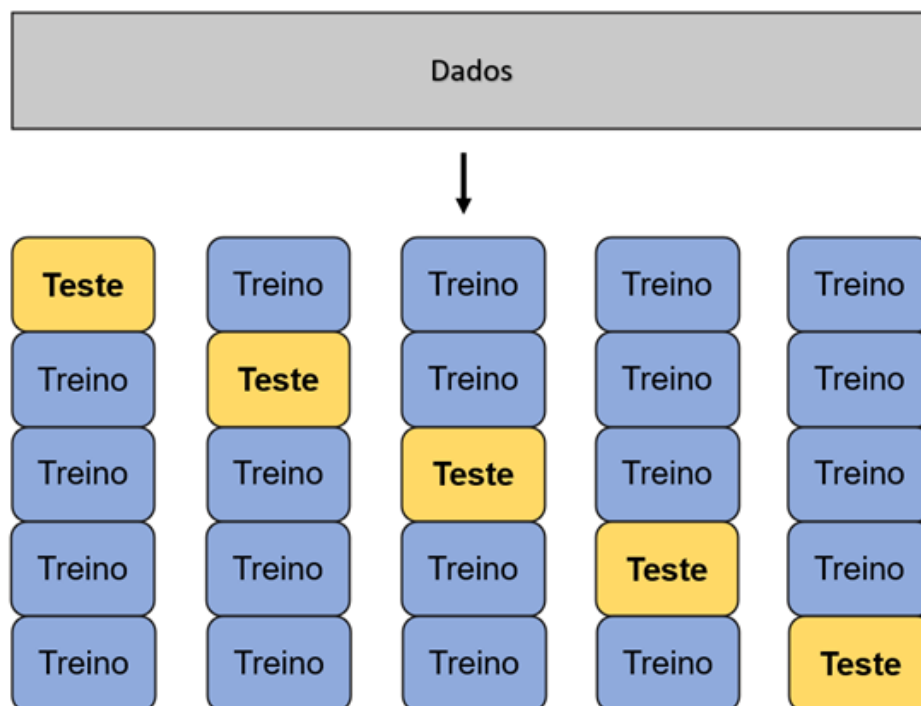
Com as extrações dos coeficientes conclui-se a etapa de pré-processamento, dando início a etapa de identificação de padrões. Para essa nova fase, foi utilizado o algoritmo 2, executável também em MATLAB. Este algoritmo foi dividido em 4 grandes etapas, elas foram: Treinamento, Adaptação de Modelos, Pontuação e Resultados. Para que a Etapa 1 fosse executada de maneira correta, uma parte das amostras contidas na base analisada foi destinada exclusivamente para treinamento, classes foram estipuladas de acordo com a análise, exemplo: Na entrega dos resultados parciais foram definidas ao total 8 classes, para a identificação dos locutores, já em resultados finais foram definidas duas classes, dor e não dor. As demais amostras

compuseram a base testes, posteriormente foi definido o número de gaussianas, sendo ele potência de 2 e variando entre 2 a 128. O número de gaussianas foi um dado muito importante utilizado na etapa 2 durante a execução do modelo, pois através dele foi indicado qual seria o refinamento durante a extração dos vetores acústicos para o posterior cálculo das probabilidades efetuado na etapa 3, onde foram feitos os testes. Por fim na 4 etapa são apresentados os resultados, com suas respectivas taxas de acertos.

4.4 VALIDAÇÃO CRUZADA

A validação cruzada é um método de reamostragem originalmente empregado para avaliar a capacidade de predição das equações de regressão linear utilizadas para prever um critério de desempenho a partir de pontuações em uma bateria de testes (BERRAR, s.d.). Esse método garante a aleatoriedade na determinação das amostras que foram destinadas exclusivamente para treinamento e teste.

Figura 6 – Ilustração do processo k -fold para $k=5$



Fonte: os Autores

Conforme ilustrado na Figura 6, na validação cruzada k -fold o conjunto de amostras disponíveis é dividido em K subconjuntos de tamanho aproximadamente igual. O modelo é treinado usando $K-1$ subconjuntos, que, juntos, representam o conjunto de treinamento (BERRAR,

s.d.). Então, o modelo é aplicado ao subconjunto restante, que é denotado como o conjunto de validação, e o desempenho é medido. O processo terá K iterações, até que cada um dos K subconjuntos tenha servido como conjunto de validação, garantindo desta forma que todas as amostras sejam utilizadas tanto para treino como para teste.

A validação cruzada será empregada durante a aplicação do algoritmo para a identificação do padrão de choro de dor e não dor, mais detalhada na seção 5.3.

5 METODOLOGIA EXPERIMENTAL

Neste capítulo serão abordados os testes iniciais, o processo de identificação do locutor e também a identificação dos sinais binários.

5.1 FAMILIARIZAÇÃO E TESTES INICIAIS

Inicialmente, para a familiarização as ferramentas citadas acima os alunos iniciaram os trabalhos realizando o reconhecimento das suas próprias vozes.

Todos gravaram 10 áudios curtos utilizando o software WAVE SURFER, compostos pela fala do próprio nome em tons sonoros diferentes, do grave ao agudo. Com os áudios gravados e salvos no formato desejado, ou seja, transformados de MP3 para WAV e posteriormente para HTK utilizando-se o algoritmo 1, foram construídas duas bases, uma para treinamento e outra para teste. Ao final, foram realizadas as simulações em MATLAB utilizando o algoritmo 2 para a validação e constatação da eficácia do reconhecimento de padrões, foram obtidos êxitos durante as simulações e a taxa de acerto foi máxima.

5.2 IDENTIFICAÇÃO DO LOCUTOR DO CHORO

Diferente dos áudios gravados pelos alunos e utilizados para a familiarização com os programas e algoritmos, as amostras fornecidas continham informações sonoras que deveriam ser descartadas, pois fugiam do padrão de qualidade mínimo esperado, ou seja, haviam áudios, com sinais de curta duração, inferiores a 1 segundo, sem som ou com vozes de outros agentes como mães e profissionais ao fundo, que ocasionavam poluições na base e queda na assertividade do reconhecimento dos locutores. Sendo assim, foi efetuada a primeira filtragem dos áudios de forma manual, identificando os trechos que continham apenas os choros dos bebês, sem ruídos e distúrbios externos e eliminando demais trechos. Neste momento as conversões de formato MP3 para WAV foram realizadas.

Ao final da primeira filtragem houve redução na base, de 19 para 16 amostras, pois não foram identificados choros nos arquivos de áudios dos bebês 3, 5 e 6, porém foi pontuado pelo especialista nos dois primeiros dados a presença da expressão de dor 8, ou seja, mesmo sem a presença do choro, os bebês estavam expressando dor através de demais sinais fisiológicos.

A base era composta exclusivamente de um áudio para cada bebê. Porém, se a mesma amostra fosse utilizada em treinamento e teste ao mesmo tempo, os resultados finais não seriam confiáveis. Por isso, com as 16 amostras disponíveis foram realizadas subdivisões, quebrando a amostra inicial em pequenas sub-amostras do mesmo bebê, ou seja, se um áudio possuísse 10 segundos de duração exclusiva do choro sem ruídos, seria dividido em 5 áudios menores de 2 segundos cada. Não houve um padrão de duração para a quebra dos áudios, sendo respeitada apenas a duração mínima estipulada de 1 segundo.

As amostras 12 ao 19 foram descartadas por não possuírem a duração mínima exclusiva de sinal de choro. Isto aconteceu devido ao ambiente onde foram coletadas as gravações estar submetido constantemente a muitas perturbações, como por exemplo ruídos oriundos de falas e aparelhos médicos.

Restaram após a segunda filtragem 8 amostras, elas foram subdivididas em 32 itens, através do fracionamento dos áudios, sendo que metade das amostras foram destinadas para treinamento e as demais para testes. Durante a experimentação, foram realizadas simulações com diferentes números de gaussianas, com o objetivo de identificar a robustez do sistema mediante a base de dados utilizada. Sendo as gaussianas, um distribuição normal, tem-se:

Para a configuração de 8 gaussianas, o resultado obtido durante a identificação foi de 87,5% de acertos, 14 casos de acertos em 16. Para a configuração de 16 gaussianas o resultado não se apresentou como esperado. Em teoria, o aumento da quantidade de gaussianas tenderia a provocar um aumento na taxa de acertos, pois mais pontos nas amostras seriam analisados. Todavia, os resultados obtidos foram uma taxa de acertos de 62,5% , ou seja, 28,57% menor se comparada a resposta com 8 gaussianas. Já com 32 gaussianas, 4 vezes a mais que a primeira simulação, o sistema apresentou a melhor resposta, com uma taxa de acerto de 93,75%, que representa uma melhora de 7,14% se comparada com a primeira simulação, sendo 15 casos de acertos em 16. Por fim, foi realizada uma última simulação com 64 Gaussianas, a resposta apresentada foi um taxa de acerto de 75%, 12 acertos em 16. Logo, analiticamente houve uma queda de 20% se comparada a simulação com 32 gaussianas, queda esperada, dada a quantidade de escassa de amostras.

A partir das identificações obtidas durante as simulações foi possível visualizar algumas particularidades deste sistema.

Observa-se que não há um padrão pré-estabelecido para a obtenção das repostas, ou seja, não é garantido que o aumento ou a diminuição do número de Gaussianas, neste caso específico, irá provocar uma resposta mais satisfatória no sistema. Isso acontece pois a base que está sendo

utilizada para o desenvolvimento deste trabalho possui poucas amostras, logo nesta primeira parte, o melhor resultado foi obtido de forma completamente experimental, através de análises sucessivas durante as simulações.

O aumento do número de gaussianas em uma base com amostras insuficientes poderá provocar a falta de dados para o cálculo de todos os parâmetros necessários para a identificação completa e correta do usuário pelo algoritmo, e há também a influência ativa do meio durante o reconhecimento, ou seja, é possível que além de reconhecer o paciente pela frequência fundamental, o algoritmo esteja também incorporando ao modelo o meio onde este bebê está inserido. Por esse motivo é de suma importância que as gravações que compõem o banco de amostras sejam realizadas no mesmo ambiente, com o mesmo tipo de dispositivo de gravação.

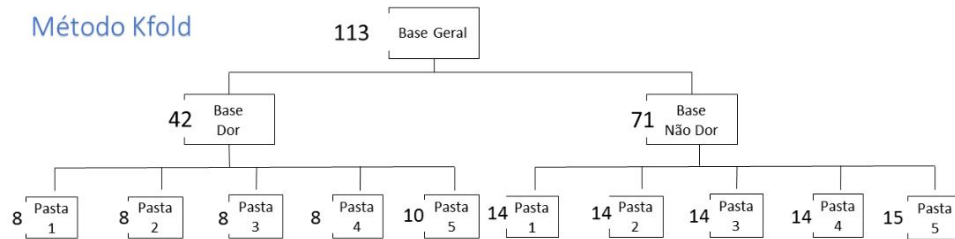
5.3 IDENTIFICAÇÃO DE PADRÕES DE DOR E NÃO DOR

Para dar segmento a segunda parte do trabalho de conclusão de curso, foi utilizada somente a base de dados do ICOPE, de maneira a uniformizar a análise, levando em consideração que, na extração dos áudios, o ruído externo também pode ser captado e influenciar na resposta à identificação, sendo o mais adequado manter o estudo com uma base captada por um mesmo equipamento e em um mesmo ambiente. Uma outra consideração importante é o potencial do algoritmo, que não é exclusivamente atrelado ao reconhecimento de locutor, uma vez que a proposta do reconhecimento de padrão do algoritmo implementado se manteve e, apenas os parâmetros utilizados foram alterados, o que na fase inicial era tido como parâmetro o reconhecimento de locutor, ou seja, a identificação da voz do choro cada bebê, foi agora modificado para parâmetros binários (dor e não dor), onde será feito um reconhecimento através das características captadas, se o choro corresponde a uma resposta de dor ou não. Como um passo inicial no processo de identificação, as amostras foram convertidas do formato .MP3 para o formato .WAV. Em seguida, executou-se o algoritmo 1 (seção 4.2) para extração dos parâmetros mel cepstrais e obtenção do arquivo HTK, que permitiu que o algoritmo 2 (seção 4.3) conseguisse identificar o padrão de cada áudio. Como mencionado na seção 4.4, foi utilizado a metodologia *k-fold* para o agrupamento das amostras, com a base composta pelos 113 áudios de duas classificações, sendo 71 áudios da categoria não dor e 42 áudios da categoria dor.

Com base nessa metodologia, foi estabelecido que os arranjos dos grupos seriam compostos por 20% para teste e 80% para o treinamento (de cada grupo), o que totalizariam 4 pastas,

com 22 amostras sendo 14 de não dor e 8 de dor, e 1 pasta contendo 25 amostras, sendo 15 de não dor e 10 de dor. A ilustração dos arranjos é exibida abaixo na Figura 7.

Figura 7 – Ilustração do processo *k-fold* para o agrupamento das amostras



Fonte: os Autores

O agrupamento dos áudios propostos permite que toda a base seja simulada, de forma que todos os áudios possam ser testados, ora passando pelo grupo de treinamento, ora sendo executado como teste.

Após a distribuição dos áudios, foram executadas simulações parametrizando o número de gaussianas igual a 2, 4, 8, 16, 32, 64 e 128, e obtiveram-se resultados conforme Quadro 6.

Quadro 6 – Quantidade de acertos de acordo com número de gaussianas

	n=2	n=4	n=8	n=16	n=32	n=64	n=128
PARCIAL DOR	23	30	30	28	25	25	23
PARCIAL N - DOR	48	49	52	50	52	51	50
TOTAL	71	79	82	78	77	76	73

Fonte: os Autores

Sob a perspectiva de porcentagem de acerto, adotando número de gaussianas igual a 2, obteve-se o pior resultado total, com 62,83% de acurácia, e os piores resultados parciais, tanto de dor, com 54,76% de acurácia, como de não dor, com 67,61% de acurácia. Aumentando exponencialmente o número de gaussianas, percebe-se um ganho porcentual expressivo até atingir-se o melhor resultado com n=8, tanto em acurácia total com 72,57%, quanto na acurácia parcial de dor, com 71,43%, e na acurácia parcial de não dor, 73,24%. Todas as relações de acurácia parciais e totais referente a cada n simulado, podem ser vistas no Quadro 7. O que se observou seguindo com o aumento de gaussianas a partir daí, foi uma queda na acurácia a cada aumento de número de gaussianas, acarretado, sobretudo, pelo comportamento dos resultados parciais de dor.

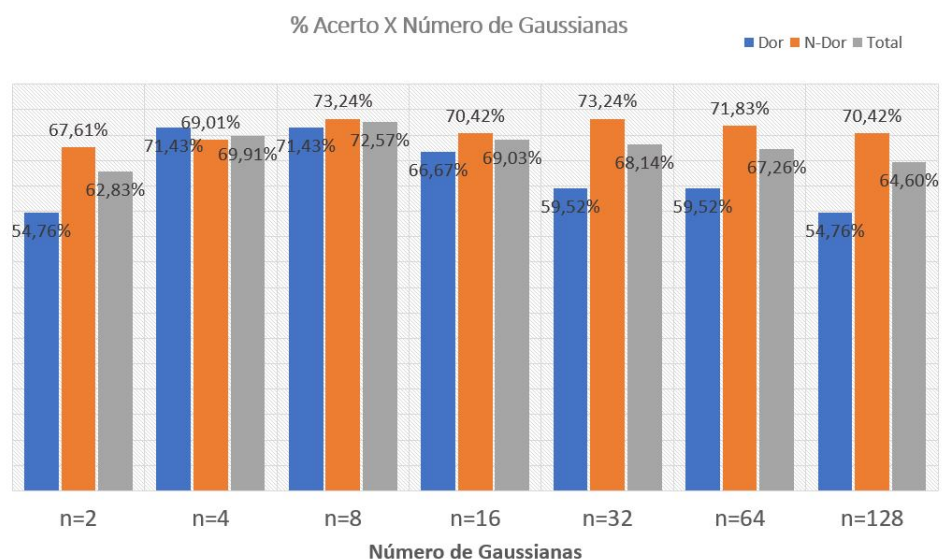
Quadro 7 – Acurácia de acordo com número de gaussianas

	n=2	n=4	n=8	n=16	n=32	n=64	n=128
PARCIAL DOR	54,76%	71,43%	71,43%	66,67%	59,52%	59,52%	54,76%
PARCIAL N - DOR	67,61%	69,01%	73,24%	70,42%	73,24%	71,83%	70,42%
TOTAL	62,83%	69,91%	72,57%	69,03%	68,14%	67,26%	64,60%

Fonte: os Autores

O aumento inicial no número de gaussianas evidenciou que, se o n for muito baixo, o parâmetro de classificação terá menos fidelidade àquilo que busca representar. No entanto, para que esse aumento do n signifique de fato uma identificação cada vez mais precisa à medida que avance, se faz necessária uma quantidade maior de dados para o algoritmo construir a característica que o mesmo deve identificar (no caso: dor e não dor). A relação entre n e a taxa de acerto apresentada pelo algoritmo pode ser verificada na Figura 8.

Figura 8 – Acurácia parcial e total versus número de gaussianas adotadas

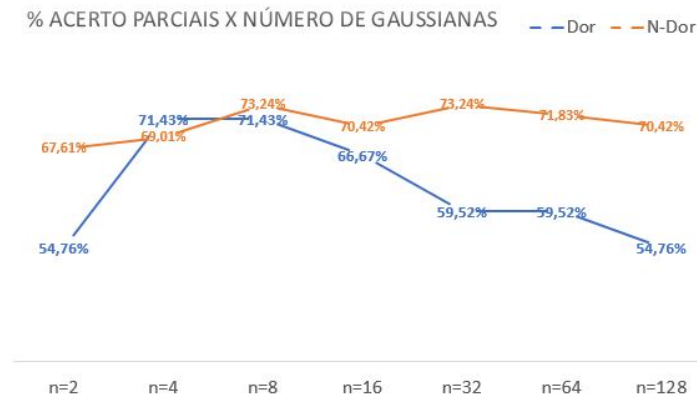


Fonte: os Autores

O desbalanceamento da base contribuiu diretamente no resultado final, e fica mais nítido, ao observar-se separadamente as acurácias de dor e não dor, conforme demonstrado em detalhes na Figura 9. Enquanto para dor a melhor acurácia é alcançada com $n=4$ e $n=8$, para a classificação não-dor a melhor acurácia é alcançada com $n=8$ e $n=32$. Ou seja, para a identificação de cada classificação, o algoritmo apresentou um comportamento diferente, tendenciando a demonstrar melhores resultados com n cada vez maiores acompanhando o tamanho da base com a classificação a ser identificada e, como a base da classificação não dor (71 áudios) é maior, tende a apresentar melhores resultados com número de gaussianas maiores, enquanto

a base da classificação de dor (42 áudios), já à partir de $n=16$ apresenta consecutivas pioras de acurácia e contribuiu diretamente para uma piora significativa nos resultados totais.

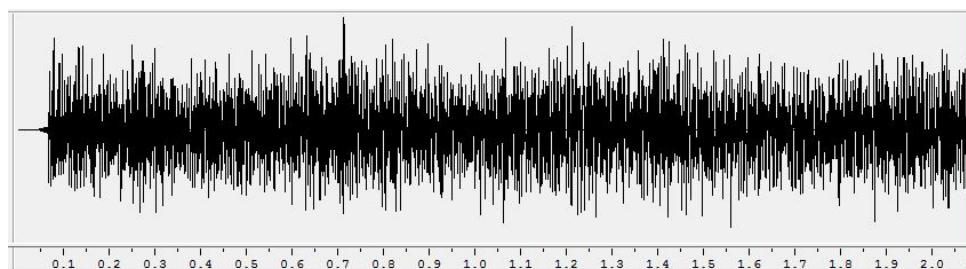
Figura 9 – Acurácia parcial (por classe) versus número de gaussianas adotados



Fonte: os Autores

Contudo, outros fatores além da variação do número de gaussianas também podem ter contribuído com os resultados obtidos, e é perfeitamente notável que há alguns outros desafios no que diz respeito à classificação automática dos choros. Através da análise empírica é possível observar que, em algumas amostras, o resultado apresentou erros para todos os números de gaussianas, o que pode ser entendido por alguns importantes fatores, sendo eles a ausência do som de choro e apenas traços de ruídos no áudio em algumas amostras e a produção da voz na inspiração, denominada fonação inspiratória. Com os áudios ruidosos, verifica-se a dificuldade na visualização e captação dos parâmetros pelo algoritmo, pois é observada uma imagem que mantém uma certa constância devido a falta de informação e a presença de ruídos, o que não se encaixa dentro da parametrização de dor ou não dor. Abaixo segue a figura 10 que exhibe o sinal no tempo da amostra citada.

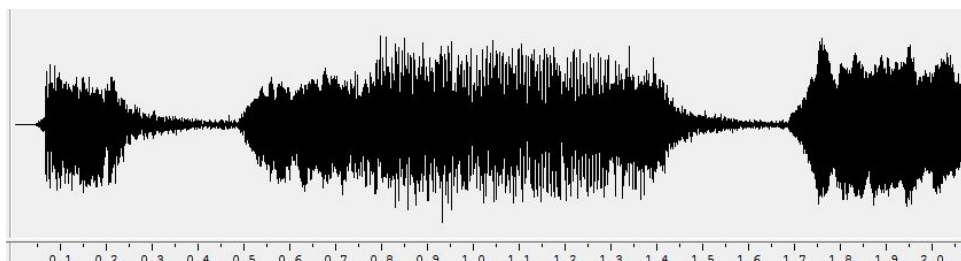
Figura 10 – Sinal no Tempo da amostra Ruidosa



Fonte: os Autores

Já na figura 11 ilustrada na sequência, é possível identificar de imediato a diferença no comportamento da onda, com extrema semelhança aos demais áudios averiguados durante o treinamento e que visualmente caracterizaria uma boa apuração pelo algoritmo.

Figura 11 – Sinal no Tempo da amostra com possível fonação inspiratória



Fonte: os Autores

A presença dessas fonações inspiratórias assemelha-se de forma extremamente significativa aos parâmetros relevantes do choro, o que pode ocasionar uma certa dificuldade para a interpretação do algoritmo e sua análise automática. Essa verossimilhança pode gerar uma confusão no que diz respeito aos limiares das unidades de choro, ocasionando algumas estimações de parâmetros espectrais e induzindo a uma resposta errônea (BAECK, 2006).

Como o algoritmo constrói o modelo das classificações embasado nos dados que são colocados à disposição, buscando a presença de áudios inicialmente categorizados como choro de dor e não-dor, e, no entanto, no Sinal no Tempo, diferem àquilo que é considerado o formato adequado à característica procurada, a consequência é a inteligência construir o modelo do que se procura combinando as formas de acordo com as amostras utilizadas. Em outras palavras, como os coeficientes mel-cepstrais são oriundos de características de frequência e busca-se a unidade do choro característica que demonstra ou não dor por parte do interlocutor, o modelo construído para as classificações em fase de treino pode conter traços de características teoricamente não presentes em um típico choro de bebê.

Em suma, essa seção apresentou análises e correlações entre parâmetros e suas acurácias respectivas e, paralelo a isso, fatores que contribuem para a "poluição" da base de dados e interferem diretamente na acurácia do resultado final. A seção seguinte trará uma abordagem mais apurada da comparação de outros estudos com esse tema.

6 COMPARAÇÃO DE METODOLOGIAS

Esta seção visa apresentar trabalhos relacionados com a base de dados do iCOPE e comparar resultados finais.

6.1 ESTUDOS RELACIONADOS

Apesar de existirem alguns estudos sobre esse tema, um estudo em específico, serviu de inspiração para realizar uma comparação de resultados finais, visto que aborda a mesma base de dados. O trabalho em questão realiza um estudo de identificação da motivação do choro de recém-nascidos utilizando recursos extraídos de um espectrograma gerado a partir de uma amostra de áudios (FELIPE et al., 2019).

O conjunto de dados utilizados para este estudo foram os mesmos dados disponibilizados pela ICOPE Database. Esses dados são separados utilizando a metodologia de validação cruzada k -fold e para classificar essas amostras, foram extraídas características diretamente do sinal de áudio, e também características de uma representação visual do sinal de áudio (FELIPE et al., 2019).

Em relação a representação visual, o primeiro passo é gerar um espectrograma a partir do áudio. O espectrograma pode ser considerado uma representação gráfica do conteúdo de áudio, na qual o eixo x representa o tempo decorrido em segundos e o eixo y representa a frequência em Hertz. Uma vez obtido o espectrograma, as características dos recursos visuais foram extraídos utilizando três diferentes descritores de textura, sendo eles: Padrão binário local, quantificação de fase local e padrão binário local robusto (FELIPE et al., 2019).

Em relação as características acústicas, os descritores acústicos utilizados foram: Escala Mel, Coeficientes Cepstrais de Frequência Mel e Cromagrama Constante-Q (FELIPE et al., 2019).

A partir da obtenção das características, estas foram utilizadas como entrada de um sistema de classificação. A classificação foi realizada utilizando a versão do algoritmo *Support Vector Machine* implementada pela biblioteca LibSVM. Esta retorna uma estimativa de probabilidade de acerto para cada classe, dor e não dor (FELIPE et al., 2019).

Mais detalhes sobre o estudo que serviu de inspiração, pode ser encontrado em (FELIPE et al., 2019).

6.2 TAXA DE ACERTO E EFICIÊNCIA

No tocante aos resultados, no trabalho (FELIPE et al., 2019) o que vale ser destacado são os resultados obtidos a partir das características acústicas, visto que esse é o enfoque da comparação de valores finais. Os resultados obtidos com os descritores acústicos Escala Mel, Coeficientes Cepstrais de Frequência Mel e Cromagrama Constante-Q, foram 64,6%, 62,83% e 62,83% respectivamente. O presente trabalho apresenta uma taxa de acurácia total de 72,57%, obtendo portanto uma porcentagem 7,97% maior.

Outros valores obtidos no trabalho (FELIPE et al., 2019) não são de grande valia a título de comparação devido ao fato de ser fusão de características visuais e características acústicas, o que acaba se tornando impossível para critério de comparação, visto que o trabalho atual só utiliza características acústicas.

7 DISCUSSÃO E CONCLUSÃO FINAL

Como pode ser observado durante o desenvolvimento desta pesquisa, a literatura referente ao estudo do choro em recém nascidos está presente no meio científico há quase um século e aponta significativos parâmetros a serem considerados na análise do sinal do choro, sendo eles: fatores anatômicos, fisiológicos e ambientais. Além disso, a literatura também pontua a extrema necessidade de que sejam ampliadas as pesquisas relacionadas a interpretação do choro pois o entendimento correto deste sinal de comunicação trará aos cuidadores e a todas as áreas de interface com recém nascidos, informações precisas para o correto e pronto atendimento.

Quanto a parte experimental, o trabalho atingiu o objetivo de validar a aplicação das ferramentas propostas, juntos os algoritmos 1 e 2 realizaram a extração e identificação dos parâmetros.

Inicialmente ambos foram desenhados com a finalidade de realizarem o reconhecimento de um parâmetro específico, o locutor, através da co-relação do grau de similaridade do áudio testado com as classes pré-estabelecidas e treinadas anteriormente, porém com a evolução deste estudo e consequentemente a ampliação do entendimento e do funcionamento dos algoritmos foi verificada a possibilidade de extensão do uso destas ferramentas, o que norteou a continuação dos desenvolvimentos propostos na entrega parcial desta monografia.

Através dos resultados obtidos foi comprovado o potencial para análise de duas classes muito distintas porém com valores próximos de amplitude e frequência, não sendo mais necessária a construção de uma nova solução para a análise de dor e não dor.

Entretanto durante a realização das simulações, reconhecimento de locutor e distinção de dor e não dor, foram observadas semelhanças na acurácia dos resultados, ambas não atingiram o valor máximo de acertos durante os testes e há diversas razões para tais desempenhos, pode-se citar: limitação na quantidade de número de amostras, poluição das amostras com a presença de ruídos com origens diversas e por fim a escassez de informações essenciais como condições físicas dos pacientes, tais déficits provocaram uma queda no desempenho do algoritmo utilizado.

A relação diretamente proporcional do número de gaussianas adotado com a quantidade de amostras treinadas e testadas também foi um dos fatores que determinou a assertividade dos testes, sendo que N maiores exigiram mais dados, para que fossem feitas análises corretas.

Além das limitações na quantidade de amostras, há outro empecilho que merece atenção: a dificuldade em se obter bases de dados de qualidade. A maioria dos bancos são particulares, restritos a uma aplicação específica e protegidos por lei.

Portanto, para estudos futuros com o objetivo de se atingir uma análise confiável, com conclusões mais precisas referentes a capacidade dos algoritmos utilizados para classificação do choro, são necessárias, além de condições padronizadas para captação do sinal de choro, bases mais robustas compostas por dados de alta qualidade e que estejam disponíveis para acesso.

REFERÊNCIAS

ASTHANA, S.; VARMA, N.; MITTAL, V. K. An investigation into classification of infant cries using modified signal processing methods. In: 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN). [S.l.: s.n.], 2015. P. 679–684. DOI: 10.1109/SPIN.2015.7095282.

BAECK, Heidi Elisabeth. **ESTUDO LONGITUDINAL DE ATRIBUTOS ACÚSTICOS DO CHORO DE BEBÊS NORMAIS (0 a10 MESES), ASSOCIADOS AO CONTEXTO DE FOME.** [S.l.], 2006.

BALDA, Rita de Cássia Xavier; GUINSBURG, Ruth. **A linguagem da dor no recém-nascido.** [S.l.], dez. 2018. Disponível em: <https://www.sbp.com.br/fileadmin/user_upload/DocCient-Neonatal-Linguagem_da_Dor_atualizDEz18.pdf>.

BERRAR, Daniel. Cross-Validation Call for Papers for Machine Learning journal: Machine Learning for Soccer View project Cross-validation. DOI: 10.1016/B978-0-12-809633-8.20349-X. Disponível em: <<https://doi.org/10.1016/B978-0-12-809633-8.20349-X>>.

BOORSTIN, Daniel J. **The Discoverers.** [S.l.: s.n.], 1983.

BRAHNAM, Sheryl et al. SVM Classification of Neonatal Facial Images of Pain. Springer, Berlin, Heidelberg, 2006. DOI: https://doi.org/10.1007/11676935_15.

BRANCO, Anete; FEKETE, Saskia Maria W.; SOUZA RUGOLO, Lígia Maria S. de. O choro como forma de comunicação de dor do recém-nascido: uma revisão. Revista Paulista de Pediatria de São Paulo, 2006. ISSN 0103-0582.

BRANCO, Anete et al. Valor e variações da frequência fundamental no choro de dor de recém-nascidos. **Revista CEFAC**, FapUNIFESP (SciELO), v. 8, p. 529–535, 4 dez. 2006. ISSN 1516-1846. DOI: 10.1590/s1516-18462006000400014.

CARDOSO, Denis Pirtiaho. **Identificação de locutor usando modelos de misturas de gaussianas.** 2009. Tese (Doutorado) – Universidade de São Paulo.

DA SILVA, Liana Albuquerque. **Dor neonatal e o Comportamento de recém-nascidos pré-termo.** [S.l.], 2017.

FELIPE, Gustavo Z. et al. Identification of Infants' Cry Motivation Using Spectrograms. In: 2019 International Conference on Systems, Signals and Image Processing (IWSSIP). [S.l.: s.n.], 2019. P. 181–186. DOI: 10.1109/IWSSIP.2019.8787318.

GARCIA, J. O.; REYES GARCIA, C. A. Mel-frequency cepstrum coefficients extraction from infant cry for classification of normal and pathological cry with feed-forward neural networks. In: PROCEEDINGS of the International Joint Conference on Neural Networks, 2003. [S.l.: s.n.], 2003. v. 4, 3140–3145 vol.4. DOI: 10.1109/IJCNN.2003.1224074.

GUINSBURG, Ruth; CUENCA, Maria Clemenza. **A linguagem da dor no recém-nascido**. [S.l.], out. 2010. Disponível em:
<https://www.sbp.com.br/fileadmin/user_upload/pdfs/doc_linguagem-da-dor-out2010.pdf>.

HANSEN, John H.L.; HASAN, Taufiq. **Speaker recognition by machines and humans: A tutorial review**. v. 32. [S.l.]: Institute of Electrical e Electronics Engineers Inc., nov. 2015. P. 74–99. DOI: 10.1109/MSP.2015.2462851.

LEDERMAN, Dror. **Automatic classification of infants' cry**. 2002. Diss. (Mestrado) – Ben-Gurion University of the Negev.

LENT, R. **Neurociência da mente e do comportamento**. [S.l.]: Guanabara Koogan, 2008. ISBN 9788527713795. Disponível em:
<<https://books.google.com.br/books?id=Zp6HPgAACAAJ>>.

MELO, Fabrício Gutemberg Lélis. **Avaliação de uso do coeficientes mel-cepstrais na representação das características vocais de um locutor**. [S.l.], 2014.

REVISTA de la Sociedad Otorrinolaringológica de Castilla y León, Cantabria y La Rioja. [S.l.: s.n.]. Disponível em: <<https://gredos.usal.es/handle/10366/124027>>.

RODRIGUES, Ernani. Reconhecimento de Voz utilizando extração de Coeficientes Mel-Cepstrais e Redes Neurais Artificiais. Instituto Federal de Santa Catarina, 2017.

RUSU, M. S. et al. Database and system design for data collection of crying related to infant's needs and diseases. In: 2015 International Conference on Speech Technology and Human-Computer Dialogue (SpeD). [S.l.: s.n.], 2015. P. 1–6. DOI: 10.1109/SPED.2015.7343077.

SANTOS, Ana Sofia Correia Dos. Sobre o choro: análise de perspectivas teóricas. **Análise Psicológica**, Instituto Superior de Psicologia Aplicada, v. 18, n. 3, p. 325–334, 2000.

SARASWATHY, J. et al. Automatic classification of infant cry: A review. In: 2012 International Conference on Biomedical Engineering (ICoBE). [S.l.: s.n.], 2012. P. 543–548. DOI: 10.1109/ICoBE.2012.6179077.

SRINIVASA, N R et al. Definição revisada de dor pela Associação Internacional para o Estudo da Dor : conceitos , desafios e compromissos, p. 1–8, 2020.

VANZELLA, Thais Peres. Normatização Dos Parametros Acústicos Vocais em Crianças em idade escolar, p. 128, 2006.

VIANA, Hesdras Oliveira. **Descritor de voz invariante ao ruído**. 2013. Diss. (Mestrado) – Universidade Federal de Pernambuco.

WASZ-HOCKERT, Ole; MICHELSSON, Katarina; LIND, John. Twenty-Five Years of Scandinavian Cry Research. In: **Infant Crying: Theoretical and Research Perspectives**. Edição: Barry M. Lester e C. F. Zachariah Boukydis. Boston, MA: Springer US, 1985. P. 83–104. ISBN 978-1-4613-2381-5. DOI: 10.1007/978-1-4613-2381-5_4.