



Universidade Federal de Itajubá

Graduação em Engenharia Eletrônica

DESENVOLVIMENTO DE UMA BABÁ ELETRÔNICA COM RECONHECIMENTO DE CHORO INFANTIL UTILIZANDO REDES NEURAIS ARTIFICIAIS

GUSTAVO S. LEAL

Orientador:

GISCARD F. C. VELOSO

Instituto de Engenharia de Sistemas e Tecnologia da Informação, Universidade Federal de Itajubá

Av. BPS 1303, 37500-903, Itajubá, MG, Brasil

E-mail: gustavo.soares.leal@gmail.com

Abstract— This article presents the development process of a infant cry monitoring device which is capable of recognizing a cry event and alert a carer remotely. The system operates by extracting mel-frequency cepstral coefficients (MFCC), which are used as characteristics for an artificial neural network-based classification. The system consists of two units: a transmitter (central) and a receptor. The transmitter is responsible for sound capturing, its digital processing and classification using artificial neural networks. The receptor is a device that receives a signal from the transmitter, triggering vibrating, sound and visual alerts.

Keywords— mfcc, neural networks, digital signal processing, cepstrum, hardware.

Resumo— Este artigo apresenta o processo de desenvolvimento de um dispositivo para monitoramento de choro infantil capaz de reconhecer um evento de choro e alertar remotamente um responsável. O sistema opera realizando a extração de coeficientes cepstrais na frequência mel (MFCC), que são utilizados como características para a classificação baseada em redes neurais artificiais. O sistema é constituído por duas unidades: um transmissor (central) e um receptor. O transmissor é responsável pela captação sonora, seu processamento digital e sua classificação utilizando redes neurais artificiais. O receptor é um dispositivo que recebe um sinal do transmissor, disparando alertas vibratórios, sonoros e visuais.

Palavras chave— mfcc, redes neurais, processamento digital de sinais, cepstrum, hardware.

1 Introdução

Segundo o censo demográfico de 2010 (Instituto Brasileiro de Geografia e Estatística 2010), no Brasil, 5,1% da população possui algum tipo de deficiência auditiva. Dois milhões de pessoas declararam ser completamente surdas (não conseguem ouvir de modo algum) ou responderam ter grande dificuldade de audição.

Em particular, o caso de pais surdos desperta a atenção quanto ao relacionamento e ao cuidado com seus bebês.

Atualmente, há dispositivos eletrônicos comercializados para monitoramento de crianças, como a chamada babá eletrônica. Este tipo de dispositivo, em sua maioria, confia na reprodução remota do som ambiente do recinto onde está uma criança. Os responsáveis são alertados quando escutam os sons emitidos. Este tipo de dispositivo, uma vez que baseia-se apenas na transmissão de informação sonora, não pode

ser usado por indivíduos surdos.

Este trabalho apresenta o desenvolvimento de um sistema embarcado capaz de detectar choro infantil, diferenciando-o de outros sons no ambiente, e enviar, remotamente, um alerta vibratório para um dispositivo remoto, funcionando assim como uma babá eletrônica com foco principal em usuários surdos ou com dificuldade de audição.

O projeto inclui o desenvolvimento de hardware e software, desde a escolha de componentes e o projeto da placa de circuito impresso, até o desenvolvimento do software, responsável por todo o processamento, que será embarcado na placa.

A figura 1 apresenta um diagrama de blocos simplificado do sistema proposto.

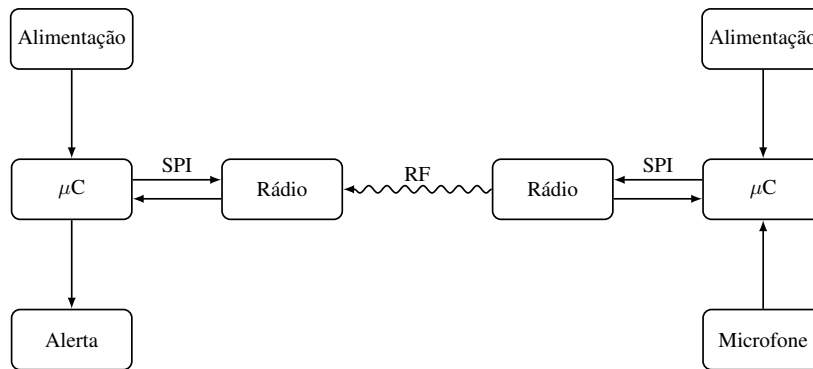


Figura 1: Diagrama de blocos do sistema proposto

2 Características do choro infantil

De acordo com (Brazelton 1962), o choro infantil é fisiologicamente útil nos primeiros dias de vida, sendo necessário para sua sobrevivência. O choro serve como ajuda na organização da função cardiopulmonar dos bebês. Ainda segundo (Brazelton 1962), logo após o nascimento, o choro começa a agir como resposta a necessidades fisiológicas, como fome ou desconforto.

(Truby e Lind 1965) classificaram como *choro básico* infantil aquele em que a criança não demonstra grande desconforto, podendo apresentar frequências fundamentais de 400 a 550 Hz e uma característica suave.

Um segundo tipo de choro apresenta característica disfônica, onde ocorrem oscilações e deformações das vias aéreas superiores, gerando um som definido como rouco ou áspero, com oscilações mensuradas em torno de 130 Hz.

Ainda, o último tipo de choro apresentado é descrito como aquele onde existe uma impressão de um som de frequência extremamente alta. Nas gravações realizadas, um choro com frequência fundamental de 400 Hz alterava-se abruptamente para 1400 Hz.

A figura 2 apresenta o conteúdo espectral de uma amostra de choro infantil.

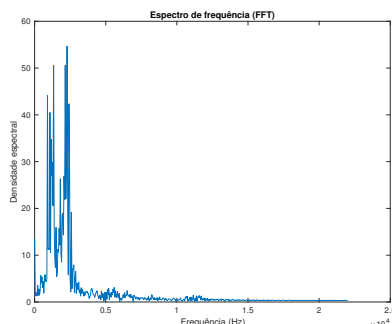


Figura 2: Conteúdo espectral de uma amostra de choro infantil

3 Audição humana e percepção psicológica de tons

3.1 Audição humana

O ouvido humano é constituído por estruturas responsáveis por captar vibrações do ar e transmiti-las ao cérebro.

No ouvido externo está localizado o canal auditivo, um pequeno tubo que se estende em direção à cabeça. Sua função é direcionar o som ambiente para os ouvidos médio e interno, que se localizam dentro do crânio. Na extremidade final do canal auditivo está o tímpano, uma fina camada que vibra quando atingida pelas ondas sonoras.

O ouvido médio constitui-se de pequenos ossos que transferem a vibração do tímpano para a cóclea, localizada no ouvido interno. (Smith 1999)

A cóclea é uma pequena estrutura espiral que possui um líquido em seu interior. O movimento do líquido coclear causado pelas vibrações deforma a membrana basilar. As características destas deformações dependem da região atingida da membrana, conferindo-a a capacidade de distinguir diferentes frequências. (Purves et al. 2004)

3.2 Escala de frequência mel

A escala de frequência mel tenta mensurar a sensação psicológica proporcionada pelo ouvido humano a cada tom.

(Stevens, Volkman e Newman 1937) construíram uma escala baseada na comparação de tons. Seu nome é oriundo da palavra *melodia*.

Para a constituição da escala, realizou-se um experimento onde os participantes ouviam um tom em uma frequência fixa e deveriam selecionar a frequência de um próximo tom de forma que este soasse como a metade do valor daquele.

A escala foi construída a partir da referência de que um tom de 1000Hz teria o valor de 1000 mels. Os demais tons foram determinados pelos participan-

tes do experimento e a fórmula utilizada para conversão (O'Shaughnessy 1987) é dada por

$$m = 2595 \log\left(1 + \frac{f}{700}\right) \quad (1)$$

e, inversamente

$$f = 700\left(10^{\frac{m}{2595}} - 1\right) \quad (2)$$

onde f é a frequência em hertz e m a frequência (ou tom) em mels. Esta relação é apresentada na figura 3.

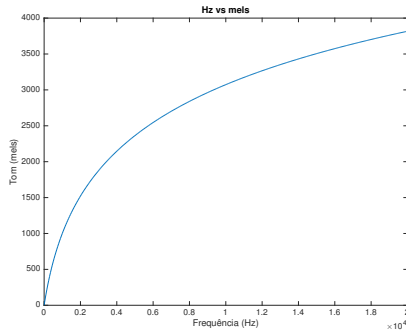


Figura 3: Relação entre escalas mel e Hz

4 Coeficientes cepstrais na frequência mel (MFCC)

Os coeficientes cepstrais na frequência mel (MFCC - Mel Frequency Cepstral Coefficients) são coeficientes obtidos a partir do *cepstrum mel*, onde, na obtenção do cepstrum, utiliza-se a escala de frequência mel. (Davis e Mermelstein 1980)

Estas características são largamente utilizadas em aplicações de reconhecimento de voz (Benesty, Sondhi e Huang 2007), mas também encontram aplicações em detecção de choro infantil (Cohen e Lavner 2012).

A implementação (Practical Cryptography 2013) para obtenção destes coeficientes, representada no diagrama de blocos da figura 5, começa pela divisão de um sinal discreto no tempo em *frames* entre 20 a 40 milissegundos de duração.

Para cada um destes *frames* (mostrados na figura 6), toma-se a transformada discreta de Fourier (DFT) das amostras, seguido pela densidade espectral, onde toma-se o valor absoluto do resultado da DFT, elevando-o ao quadrado e dividindo pelo número de amostras do sinal.

O resultado passa por um banco de filtros passa-banda - como o mostrado na figura 4 - gerados a partir da conversão de mel em hertz. Esta divisão é feita para avaliar padrões de distribuição de energia ao longo do espectro.

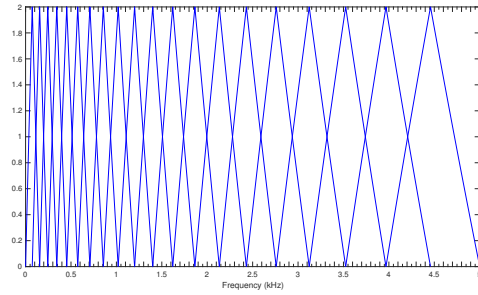


Figura 4: Banco de filtros passa-banda

Soma-se, então, o resultado da aplicação de cada um destes filtros (chamado *energia*) individualmente. Toma-se o logaritmo das energias acima obtidas, seguido pela transformada cosseno discreta (DCT). Os valores resultantes são os coeficientes desejados. É comum, na maioria das aplicações, manter os 13 primeiros coeficientes, por terem maior informação na faixa de frequência da voz humana. Neste trabalho, portanto, estes 13 coeficientes foram utilizados como entrada para a rede neural. (Practical Cryptography 2013)

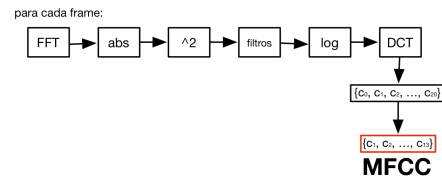


Figura 5: Diagrama de blocos do algoritmo para extração dos coeficientes

Para a aplicação proposta, utilizou-se uma frequência de amostragem de 10 kHz. O conversor analógico digital adquire 1280 amostras e as armazena em um vetor, posteriormente dividindo-o em *frames* de 25,6 ms de duração (correspondentes a 256 amostras) com 64 amostras sobrepostas. A sobreposição de amostras é feita para diminuir a perda de informações causadas pela aplicação de uma janela antes do processamento da FFT.

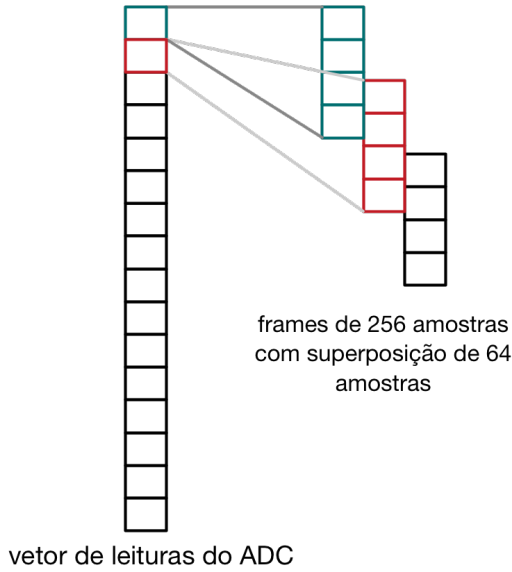


Figura 6: Representação da divisão em frames

As operações foram efetuadas utilizando aritmética de ponto fixo no formato Q1.15, com um bit decimal e 15 bits fracionários. A conversão de um valor *float* para um valor Q1.15 se dá pelo arredondamento do resultado da multiplicação do valor em ponto flutuante por 32768 (uma vez que, para variáveis sinalizadas de 16 bits, os valores variam entre -32768 e 32767).

Utilizou-se a *toolbox* VOICEBOX (Brookes 2003), para a geração do banco de filtros. A implementação define o número de filtros por $\lceil 4.6 * \log(f_s) \rceil$, onde f_s é a frequência de amostragem de 10 kHz, resultando em 19 filtros, compreendendo a faixa de 0 a 5 kHz (0 a 2363,46 mel).

5 Redes neurais e reconhecimento de padrões

Para classificação do choro infantil, optou-se por utilizar redes neurais artificiais. Embora requeiram maior esforço computacional, os tamanhos das matrizes de pesos e vieses de uma rede neural dependem apenas da quantidade de entradas e de neurônios. Em uma implementação inicial, utilizou-se um classificador baseado em regras *if-then*, como uma árvore de decisões. A aplicação deste tipo de algoritmo para este trabalho, embora computacionalmente mais leve, mostrou-se inviável, uma vez que o tamanho do código aumentava quanto mais amostras eram utilizadas para o treinamento.

Como o sistema conta com um microcontrolador com recursos de memória bastante limitados, um equilíbrio foi encontrado entre uso de processamento e me-

mória na aplicação de redes neurais.

Redes neurais são estruturas baseadas em um modelo de um neurônio real, constituído de múltiplas entradas e uma saída. Cada entrada é multiplicada por um **peso**. Estas entradas, então, são comparadas a um valor de limiar e uma **função de ativação**, que determina sua saída.

Aqui, utilizou-se uma estrutura chamada **rede de retropropagação**, sendo a mais utilizada para implementações de redes neurais artificiais. Esta construção consiste num conjunto de neurônios formando uma **camada**. Todas as entradas de uma camada são conectadas à camada anterior ou às entradas da rede neural. Da mesma forma, todas as saídas de uma camada são conectadas à próxima camada ou à saída.

Entre a camada de entrada e a camada de saída pode haver camadas intermediárias, chamadas **camadas ocultas** por não possuírem ligação com as variáveis externas à rede.

Uma rede neural com retropropagação é representada pelo seguinte diagrama (figura 7):

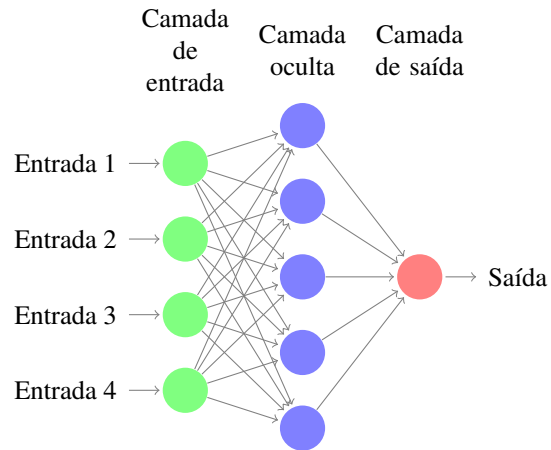


Figura 7: Rede neural com uma camada oculta

A saída de um neurônio U_j , sendo j sua ordem, é relacionada à sua entrada X_i , sendo i a ordem da entrada, pela função

$$U_j = \sum (X_i * w_{ij}) \quad (3)$$

onde w é o peso pré-estabelecido. (Cheshire Engineering Corporation 1994)

O resultado desta operação (U_j) é somado a um valor chamado **viés** e encaminhado para uma **função de ativação**. Sendo Y_j a saída do neurônio, b_j o valor do viés para o neurônio j e F uma função de ativação, tem-se que

$$Y_j = F(U_j + b_j) \quad (4)$$

As funções de ativação são funções não-lineares, o que possibilita às redes neurais realizar um mapeamento não-linear entre as entradas e saídas, próximo

ao comportamento de um neurônio real. Ainda, são funções limitadas superior e inferiormente, garantindo que, independente da entrada, seus valores estarão sempre entre seus limites. Vas 1999

A função de ativação é, comumente, uma função sigmoide, como $f(x) = \frac{1}{1+e^{-x}}$ (representada na figura 8) ou ainda a função tangente hiperbólica $f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$. (Multilayer Perceptron)

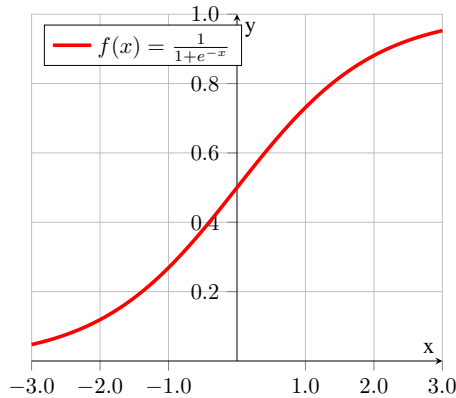


Figura 8: Função de ativação sigmoide

Para o treinamento da rede neural aplicada, utilizou-se o pacote *Neural Network Toolbox* do *MATLAB*®.

Como entrada, utilizou-se uma matriz cujas linhas correspondiam aos vetores contendo os 13 coeficientes calculados para cada *frame*.

Para indicar o resultado, foram atribuídas duas classes: 0 para *não houve choro* e 1 para *houve choro*.

Baseado em (García e García 2003), o treinamento foi iniciado utilizando o algoritmo *scaled conjugate gradient*. As amostras foram divididas em 60% para treino, 10% para validação e 30% para teste. O diagrama na figura 9 mostra a configuração da rede gerada

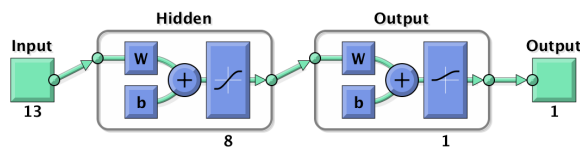


Figura 9: Diagrama gerado pelo *MATLAB*® de uma rede neural como a utilizada na aplicação

6 Implementação

6.1 Projeto de hardware

Para a placa que realiza a captura e o processamento dos dados (Central), são requerimentos:

- Microfone para captura de áudio

- Conversão analógico-digital da leitura de um sinal de microfone
- LEDs para indicação de estado
- Módulo de radiofrequência para envio de alerta à outra placa

Para a placa receptora, os requerimentos são:

- Tamanho reduzido para permitir uso como pulseira
- Motor de vibração para emissão de alerta vibratório
- *Buzzer* para emissão de alerta sonoro
- LEDs para indicação de estado
- Módulo de radiofrequência para recepção do sinal de alerta
- Baixo consumo de energia para permitir a operação com baterias

6.1.1 Bloco de processamento

Central Para a aquisição e processamento, escolheu-se um microcontrolador NXP Kinetis KL25Z com núcleo ARM® Cortex-M0+, buscando equilibrar desempenho e custo. Ainda, devido à larga utilização da plataforma ARM, os fabricantes oferecem diversos exemplos de uso, além de ferramentas para facilitar a configuração e a programação, diminuindo o tempo de desenvolvimento.

O programa é desenvolvido em linguagem C e faz uso das bibliotecas ARM® Cortex® CMSIS para as funções de processamento digital de sinais e do Kinetis Software Development Kit, constituído de drivers otimizados para o modelo de microcontrolador escolhido, acelerando o tempo de desenvolvimento e o desempenho das tarefas executadas.

Receptor Para o receptor, escolheu-se o microcontrolador Atmel ATmega328P por tratar-se de um modelo popular, utilizado em placas de desenvolvimento de hardware aberto como Arduino® UNO, possuindo assim, grande quantidade de recursos de software e documentação.

Decidiu-se pela utilização do bootloader do Arduino, que facilita a gravação do microcontrolador e permite aproveitar bibliotecas já existentes para a plataforma, agilizando o desenvolvimento.

6.1.2 Bloco de comunicação

Para a transmissão de dados, foi escolhido o transceptor Texas Instruments CC1101 operando na frequência de 915MHz. Este transceptor se comunica com um microcontrolador através de um protocolo serial síncrono (SPI - Serial Peripheral Interface).

Nas duas placas, utilizou-se o design de referência da fabricante (Texas Instruments 2010) para assegurar

o desempenho em termos de comunicação e compatibilidade eletromagnética.

Considerando a aplicação, não foi necessário implementar protocolos de comunicação que envolvem redes com mais de dois elementos, como ZigBee (IEEE802.15.4) ou Wi-Fi (IEEE802.11). Assim, foi possível reduzir a complexidade do código utilizando o transceptor com um protocolo de transmissão proprietário simples que se aplica ao formato do pacote utilizado pelo transceptor. O formato do pacote é mostrado na figura 10.

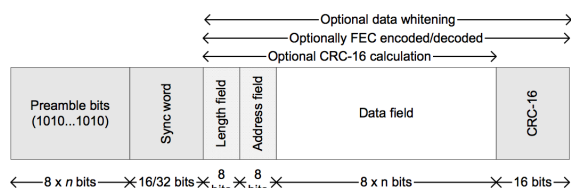


Figura 10: Formato do pacote utilizado pelo transceptor CC1101

O transceptor possui, inerentemente, suporte a mecanismos de detecção de erros e segurança, utilizando CRC. Ainda, é possível atribuir um endereço individual a cada sistema, evitando interferências no caso de proximidade.

Central De modo a reduzir o tamanho e o custo de produção da placa, escolheu-se utilizar uma antena feita com trilhas da própria placa de circuito impresso.

A fabricante do transceptor oferece um guia com modelos para antenas de diferentes tipos, aplicações e frequências. (Wallace e Texas Instruments 2010)

Aqui, para a frequência de 915MHz, utilizou-se uma antena monopolo (*Meandering Monopole*). Se houver espaço suficiente na placa de circuito impresso, o fabricante recomenda este tipo de antena como primeira opção, uma vez que apresenta o melhor desempenho dentre as alternativas. (Wallace e Texas Instruments 2010)

6.1.3 Bloco de alimentação

Central O sistema deve operar utilizando uma fonte de tensão externa comum.

Como o microcontrolador e os periféricos operam com uma tensão de 3,3V, foi incluído um regulador de tensão de baixa perda (LD1117-3.3) para fornecer e estabilizar a tensão. Este regulador pode operar com tensões de entrada de até 15V (STMicroelectronics 2005), mas, para evitar aquecimento excessivo, é ideal utilizar uma fonte externa de 5V.

Este bloco conta, ainda, com uma chave liga-desliga, interrompendo ou permitindo a conexão ao terminal de alimentação positivo.

Como proteção, foram inseridos diodos Schottky para evitar danos em caso onde, por exemplo, os polos da fonte sejam invertidos.

6.1.4 Bloco de captura de som

Decidiu-se pela utilização de um microfone do tipo MEMS (*Microelectromechanical systems*) que ocupa menor espaço na placa. O modelo utiliza um resistor externo para determinar o ganho, limitado em 20dB, e um capacitor para definir sua frequência de corte.

6.1.5 Bloco de alerta

O bloco de alerta presente no receptor é constituído por um LED, um motor vibratório e um *buzzer*. O motor e o buzzer são ligados a portas digitais do microcontrolador capazes de externar um sinal PWM, o que permite controlar a intensidade da vibração e especificar a frequência do som emitido pelo buzzer.

6.2 Projeto de software

Para desenvolvimento de software, utilizou-se das bibliotecas CMSIS-DSP, que implementam rotinas de processamento digital de sinais em linguagem C compatíveis com vários processadores de arquitetura ARM.

Para compatibilidade de hardware, utilizou-se o Kinetis Software Development Kit v2.0, fornecido pela fabricante do microcontrolador, que fornece drivers otimizados e facilita o início do desenvolvimento.

O software consiste de uma aplicação *bare-metal*, dispensando o uso de sistema operacional.

O diagrama a seguir (figura 11) mostra o fluxo da aplicação executada pelo microcontrolador.

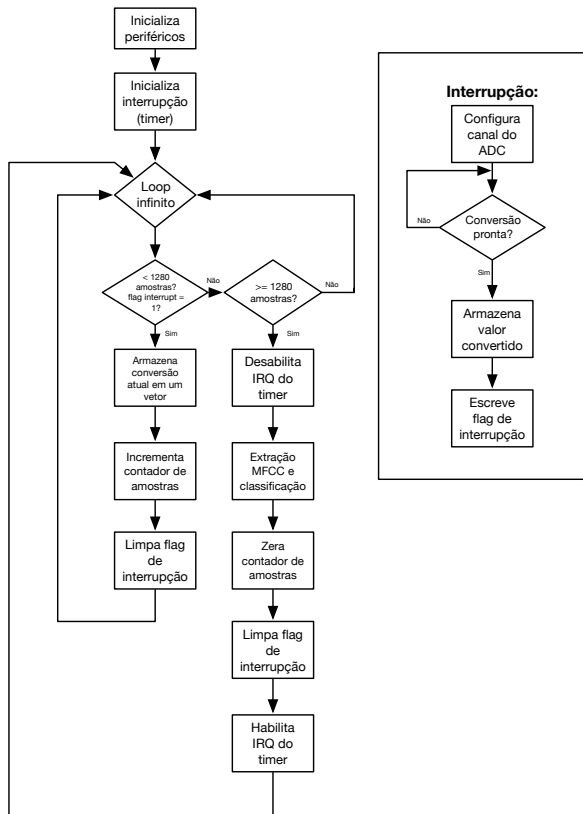


Figura 11: Fluxo da aplicação implementada

7 Resultados

7.1 Testes de captura e processamento

O sistema foi programado com a aplicação apresentada na figura 11.

Executaram-se testes onde, exposto a uma fonte sonora que reproduzia amostras de choro, o protótipo enviava ao computador, via conversor serial-USB externo, os coeficientes calculados.

Para verificar a integridade da captura das amostras, também foram executados testes onde enviava-se, via serial, o vetor com os valores capturados diretamente pelo conversor analógico-digital, dados que permitiram a visualização do sinal capturado no domínio do tempo.

Nestes testes, utilizou-se o chaveamento de um LED para, com o auxílio de um osciloscópio digital, determinar o tempo de duração de cada parte do programa.

A rotina que captura as 1280 amostras e as armazena em um vetor durou 80ms, enquanto o processamento para extração das características levou 15ms.

7.2 Reconhecimento de choro

O treinamento da rede neural utilizando os coeficientes obtidos utilizou 45580 *frames* de choro infantil

para entrada (amostras positivas, às quais foi atribuída a classe 1), e 201019 *frames* contendo ruídos diversos, como ruídos de construção civil, movimento de automóveis, sirenes e vozes humanas sobrepostas (amostras negativas, às quais atribuiu-se a classe 0).

Para determinar o número de neurônios ocultos, repetiu-se o treinamento em várias iterações, variando de 2 a 20 neurônios.

7.3 Discussão dos resultados

Como medida de desempenho do sistema, utilizou-se a interpretação de uma matriz de confusão.

Uma matriz de confusão apresenta o desempenho de um classificador em uma tabela simples:

n = 190	Previsto 0	Previsto 1	
Verdadeiro 0	60	25	
Verdadeiro 1	15	90	150

Tabela 1: Exemplo de matriz de confusão

A tabela 1 mostra um exemplo de matriz de confusão para uma classificação binária de 190 entradas. Destas, 60 entradas que pertenciam à classe 0 foram corretamente classificadas e 25 entradas que pertenciam à classe 0 foram incorretamente classificadas como sendo da classe 1, 15 entradas da classe 1 foram incorretamente classificadas como sendo da classe 0 e 90 entradas da classe 1 foram corretamente classificadas. A última célula mostra o valor total de acertos.

Após o treinamento realizado com diferentes números de neurônios, as matrizes de confusão dos melhores resultados obtidos são apresentados nas figuras 12, 13 e 14.

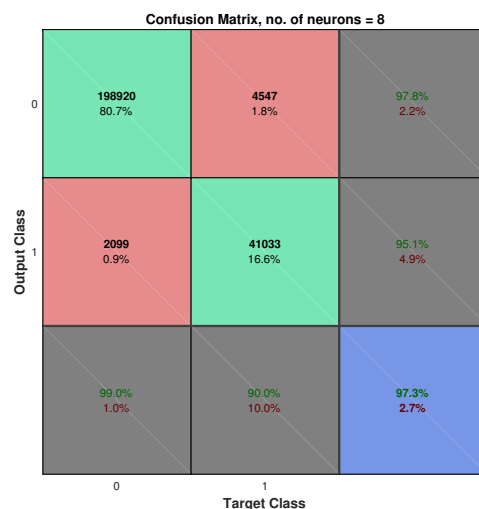


Figura 12: Matriz de confusão para 8 neurônios ocultos

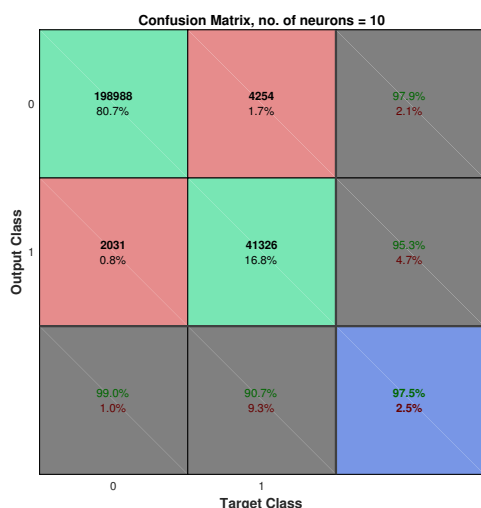


Figura 13: Matriz de confusão para 10 neurônios ocultos

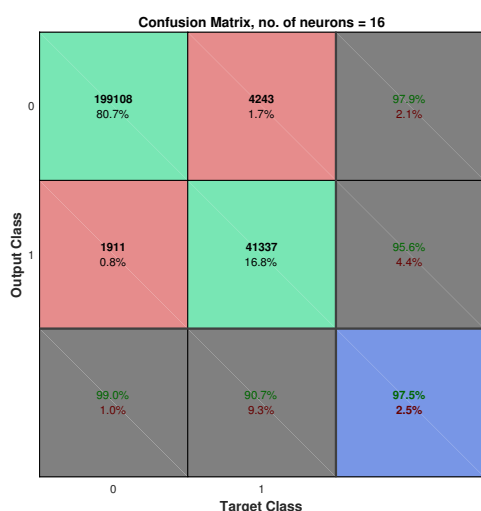


Figura 14: Matriz de confusão para 16 neurônios ocultos

Como a implementação do classificador é feita num sistema com memória limitada, é coerente escolher o melhor resultado com menor número de neurônios.

Na figura 12, as primeiras duas células diagonais da matriz de confusão indicam que 198920 amostras, correspondentes a 80,7% do total, foram classificadas corretamente como 0 (não houve choro) e 41033 amostras, 16,6% do total, foram classificadas corretamente como 1 (houve choro). Ainda, indica-se que 4547 amostras (1,8% do total) foram incorretamente classificadas como 0 (*não houve choro*) e 2099 (0,9% do total) foram classificadas incorretamente como 1

(*houve choro*).

A figura 12 ainda indica que, de 203467 predições 0 (soma das duas primeiras células, 198920 + 4547), 97,8% foram classificadas corretamente. Na linha abaixo, indica-se que, de 43132 predições 1 (2099 + 41033), 95,1% foram corretamente classificadas. A última célula da primeira coluna indica que 99% de 201019 (soma das duas células acima) casos 0 foram corretamente classificados. A última célula da segunda coluna indica que, de 45580 casos 1, 90% foram classificados corretamente.

Finalmente, a última célula da matriz da figura 12 mostra que, no geral, 97,3% das previsões foram corretas.

8 Conclusão e trabalhos futuros

Este trabalho apresentou a possibilidade de implementação de um sistema de reconhecimento de choro infantil em um microcontrolador de baixo custo e potência.

A utilização de redes neurais para classificação permitiu que a detecção funcionasse mesmo em casos onde algum fator variava, como a distância entre a criança e o microfone.

Ainda, a extração dos coeficientes cepstrais na frequência mel permitiu melhor caracterização das vocalizações produzidas durante o choro. Este método pode ser estendido para detectar além da ocorrência do choro, a sua motivação. Se forem obtidas amostras ligadas a reações da criança, como fome, dor ou desconforto, pode-se treinar a rede neural para identificar diferenças entre os choros que expressam cada sensação, o que pode ser útil no tratamento com a criança.

Ademais, o funcionamento do receptor pode ser estendido para receber sinais de outros dispositivos, que podem ser integrados em uma rede local utilizando protocolos de rede para aplicações residenciais.

Agradecimentos

O autor agradece à empresa Incluir Tecnologia pela permissão de realizar e abordar o desenvolvimento do projeto.

Referências

- Benesty, J., M. Sondhi e Y. Huang (2007). *Springer Handbook of Speech Processing*. Springer Handbook of Speech Processing. Springer Berlin Heidelberg. ISBN: 9783540491255.
- Brazelton, T. B. (1962). "Crying in Infancy". Em: *Journal of the American Academy of Pediatrics*.

- Brookes, M. (2003). *VOICEBOX: Speech Processing Toolbox for MATLAB*. URL: <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>.
- Cheshire Engineering Corporation (1994). *Neuralyst User's Guide*. URL: <https://www.cheshireeng.com/Neuralyst/doc/NUG14x.pdf>.
- Cohen, R. e Y. Lavner (2012). "Infant Cry Analysis and Detection". Em: *2012 IEEE 27th Convention of Electrical & Electronics Engineers in Israel (IEEEI)*. IEEE.
- Davis, S. B. e P. Mermelstein (1980). "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences". Em: *IEEE Transactions on acoustics, speech, and signal processing, vol. ASSP-28, no. 4*. Ed. por Wiley.
- García, J. O. e C. A. R. García (2003). "Mel-Frequency Cepstrum Coefficients Extraction from Infant Cry for Classification of Normal and Pathological Cry with Feed-forward Neural Networks". Em: *Proceedings of the International Joint Conference on Neural Networks, 2003*.
- Instituto Brasileiro de Geografia e Estatística (2010). *Censo Demográfico 2010 - Características gerais da população, religião e pessoas com deficiência*. URL: http://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd_2010_religiao_deficiencia.pdf.
- LISA lab. *Multilayer Perceptron*. URL: <http://deeplearning.net/tutorial/mlp.html>.
- O'Shaughnessy, D. (1987). *Speech communication: human and machine*. Addison-Wesley Publishing Company. ISBN: 0-2011652-0-1.
- Practical Cryptography (2013). *Mel Frequency Cepstral Coefficient (MFCC) tutorial*. URL: <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>.
- Purves, D. et al. (2004). *Neuroscience*. 3ª ed. Sinauer Associates, pp. 290–291. ISBN: 0878937269.
- Smith, S. W. (1999). *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Publishing. ISBN: 0-9660176-6-8.
- Stevens, S. S., J. Volkman e E. B. Newman (1937). "A Scale for the Measurement of the Psychological Magnitude Pitch". Em: *The Journal of the Acoustical Society of America*.
- STMicroelectronics (2005). *LD1117 series low drop fixed and adjustable positive voltage regulators datasheet*. URL: <https://www.sparkfun.com/datasheets/Components/LD1117V33.pdf>.
- Texas Instruments (2010). *CC1101EM868-915MHz Reference Design 3.0 (Rev. B)*. URL: <http://www.ti.com/lit/zip/swrr045>.
- Truby, H. M. e J. Lind (1965). "Cry sounds of the newborn infant". Em:
- Vas, P. (1999). *Artificial-Intelligence-based Electrical Machines and Drives: Application of Fuzzy, Neural, Fuzzy-neural, and Genetic-algorithm-based Techniques*. Monographs in Electrical and Electronic Engineering. OUP Oxford.
- Wallace, R. e Texas Instruments (2010). *Application Note AN058 - Antenna Selection Guide*. URL: <http://www.ti.com/lit/an/swra161b/swra161b.pdf>.