

Lucas Ferreira Paiva

**Estimação do compasso musical do forró
utilizando rede perceptron multicamadas**

Viçosa, MG

2020

Lucas Ferreira Paiva

Estimação do compasso musical do forró utilizando rede perceptron multicamadas

Monografia apresentada ao Departamento de Engenharia Elétrica do Centro de Ciências Exatas e Tecnológicas da Universidade Federal de Viçosa, para a obtenção dos créditos da disciplina ELT 402 – Projeto de Engenharia II – e cumprimento do requisito parcial para obtenção do grau de Bacharel em Engenharia Elétrica.

Orientador Rodolpho Vilela Alves Neves

Viçosa, MG

2020

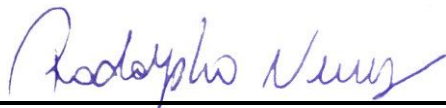
LUCAS FERREIRA PAIVA

**ESTIMAÇÃO DO COMPASSO MUSICAL DO FORRÓ
UTILIZANDO REDE PERCEPTRON MULTICAMADAS**

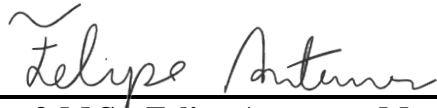
Monografia apresentada ao Departamento de Engenharia Elétrica do Centro de Ciências Exatas e Tecnológicas da Universidade Federal de Viçosa, para a obtenção dos créditos da disciplina ELT 402 – Projeto de Engenharia II e cumprimento do requisito parcial para obtenção do grau de Bacharel em Engenharia Elétrica.

Aprovada em 16 de julho de 2020.

COMISSÃO EXAMINADORA



Prof. Dr. Rodolpho Vilela Alves Neves - Orientador
Universidade Federal de Viçosa



Prof. M.Sc. Felipe Antunes - Membro
Instituto Federal de Minas Gerais – Campus Ipatinga



M.Sc. Kevin Braathen de Carvalho - Membro
Universidade Federal de Viçosa

*Este trabalho é dedicado à todas as pessoas amantes do forró,
Surdos e ouvintes.*

Agradecimentos

Agradeço primeiramente minha mãe, Gisléia, pelo exemplo de força, carinho e suporte durante toda minha vida. Ao meu pai, José Antônio e minha irmã Aline por todo apoio e incentivo. Também a toda minha família, avós e avôs, tias e tios, primas e primos, por toda fé, força e esperança confiadas a mim. Em especial, minha “família da quarentena”, Sâmara, Elizabeth e Rubens, por todo carinho e apoio.

Agradeço a todos os amigos espirituais que seguiram me fortalecendo, protegendo e intuindo em toda essa jornada de aprendizagem técnica e científica, mas também espiritual e de autoconhecimento.

Ao professor e orientador Rodolpho Vilela, por aceitar prontamente orientar esse trabalho, agradeço por toda motivação, paciência, conhecimento e tempo dedicado.

Ao professor Carlos Sperber, por me acolher nos projetos de pesquisa em ensino em um dos períodos mais difíceis da minha graduação, seus conselhos e ensinamentos ao longo dessa parceria me tornou uma pessoa mais crítica e capacitada para a vida acadêmica e pessoal.

Ao professor Mauro Prates, pela orientação do estágio no “Curso Introdutório de Eletricidade Básica e Instalações Elétricas Residenciais” pela generosidade e empatia.

Agradeço a todos os amigos da ELT em especial ao “Bondinho da Barra E”, por todo companheirismo, auxílio e parceria. Obrigado pelos momentos de estudo, diversão e de lamentação. Sozinho eu nunca teria conseguido o apoio de vocês foi essencial para que eu chega-se até aqui.

Agradeço aos companheiros do Alojamento 1721, da Religare, da Tribo do Morro, do Projeto Sementes do Amanhã, do GEDAE e do NIAS pela amizade e auxílio no crescimento profissional, acadêmico, pessoal e espiritual.

Agradeço a Priscila por ter dançado todas as 82 músicas utilizadas nesse trabalho, sua contribuição viabilizou essa pesquisa. Ao Hugo que está implementando o aplicativo, muito obrigado por ajudar a materializar esse sonho. Agradeço aos projetos “Projeto Pé de Serra no Campus” e “Dança nas Moradias” pelo fornecimento do espaço, apoio e sugestões.

Finalmente, agradeço aos professores e funcionários do departamento, por todo conhecimento e suporte ao longo do curso. Agradeço também aos demais funcionários, da UFV e terceirizados, por toda estrutura oferecida, pela comida quentinha, pela limpeza da UFV e a manutenção dos prédios e *sites*, muito obrigado, a dedicação de vocês é essencial para este e todos os trabalho feitos na UFV.

*“É pra surdo ouvir, pra cego ver
Que este xote faz milagre acontecer”.
(Xote dos milagres, Falamansa)*

Resumo

Por meio da Lei N° 13.409 de 2016, que altera a “Lei das Cotas” (Lei N° 12.711 de 2012), tem-se a reserva de vagas para pessoas surdas no ensino superior, garantindo maior acesso desses estudantes à universidade. Ao mesmo tempo, há uma forte presença do forró na cultura universitária. O forró é marcado por instrumentos de percussão, como o triângulo e a zabumba, e por compassos musicais que ditam a velocidade de movimento da dança. Visando a inclusão do público surdo em atividades culturais envolvendo o forró, este trabalho propõe estimar o tempo de compasso musical em músicas de forró utilizando uma rede perceptron multicamadas (PMC). Um banco de dados com músicas de Forró Pé-de-serra e Universitário foi utilizado para treinamento da PMC. As saídas desejadas foram obtidas com o auxílio de uma instrutora de forró, que mediu o tempo de execução dos passos da dança para cada música. A estimação do tempo de compasso realizada pela PMC atingiu erro médio percentual aproximado de 4%, permitindo que pessoas surdas consigam sentir a música com erros médios de ± 60 ms. A aplicabilidade da abordagem será tratada em trabalhos futuros.

Palavras-chaves: Rede perceptron multicamadas; Estimação de parâmetros; Inclusão de surdos; Forró; Dança.

Abstract

By the Brazilian Law No. 13,409 of 2016, that changes the “Law of Quotas” (Brazilian Law No. 12,711 of 2012), there is a reserve of quotas for deaf people in the higher education, ensuring greater access for these students in the universities. At the same time, inside the university culture, there is a strong cultural presence of “farró”, an activity to gather students with their peers. The “farró” is a musical genre marked by percussion instruments, such as triangle and zabumba, and by musical compass that dictates the movement speed of the dance. Aiming the inclusion of deaf people in cultural activities involving “farró”, this paper proposes to estimate the musical compass time for “farró” musics using a multilayer perceptron neural network (MLP). A “farró” music database with “Farró Pé-de-serra” and “Farró Universitário” is used to train the MLP. The desired outputs used to train the MLP are the musical compass time. The musical compass times were obtained for the musics in the database by a “farró” instructor, that danced the songs and measured the time to perform the fundamental steps. The musical compass times estimated by the MLP reached percent mean error of 4%, allowing deaf people to feel the music with a mean error of ± 60 ms. The applicability of the approach will be addressed in future works.

Key-words: Multilayer perceptron; Parameter estimation; Deaf inclusion; Farró; Dance.

Lista de ilustrações

Figura 1 – Modelo implementado para a estimação do ritmo da música pelo compasso musical.	13
Figura 2 – Diagrama do Passo Base Frente e Trás. O pé esquerdo (E) e direito (D) são representados pelos quadrados com as letras e as cores azul escuro e azul claro representam o pé de sustentação e o pé que toca o chão, respectivamente.	16
Figura 3 – Estrutura de uma rede PCM com uma camada oculta.	18
Figura 4 – Espaço de dança “Forró do Itaú” no dia da gravação das músicas. . . .	20
Figura 5 – Processo de seleção das melhores redes para a estimação do compasso de músicas de forró para as músicas em MP3 e WAV.	27
Figura 6 – EPM das redes PMC, para as músicas em MP3, em função do número de neurônios na camada oculta.	28
Figura 7 – Saída estimada para os dados de teste pela rede com o menor EPM (Rede 1)	29
Figura 8 – Erro obtido com os dados de teste pela rede com o menor EPM (Rede 1), referente à Figura 7.	30
Figura 9 – EPM das redes PMC, para as músicas em WAV, em função do número de neurônios na camada oculta.	31
Figura 10 – Saída estimada para os dados de teste pela rede com o menor EPM (Rede 2).	31
Figura 11 – Erro obtido com os dados de teste pela rede com o menor EPM (Rede 2), referente à Figura 10.	32
Figura 12 – EPM das redes PMC, para o banco misto, em função do número de neurônios na camada oculta.	33
Figura 13 – Saída estimada para os dados de teste pela rede com o menor EPM (Rede 3).	33
Figura 14 – Erro obtido com os dados de teste pela rede com o menor EPM (Rede 3), referente à Figura 13.	34
Figura 15 – Desempenho das Redes 2 e 3 no Teste 2 (Banco em MP3): A) Valores estimados pela Rede 2. B) Erro percentual da Rede 2. C) Valores estimados pela Rede 3. D) Erro percentual da Rede 3.	35
Figura 16 – Desempenho das Redes 1 e 3 no Teste 3 (Banco em WAV): A) Valores estimados pela Rede 1. B) Erro percentual da Rede 1. C) Valores estimados pela Rede 3. D) Erro percentual da Rede 3.	36

Lista de tabelas

Tabela 1 – Músicas do banco de dados com o interprete da versão utilizada e a duração do compasso.	21
Tabela 2 – Resumo dos resultados obtidos no Teste 1.	32
Tabela 3 – Resumo dos resultados obtidos no Testes 2 e 3.	37

Lista de abreviaturas e siglas

DP	Desvio Padrão
EP	Erro Percentual
EPM	Erro Percentual Médio
EQM	Erro Quadrático Médio
FFT	Fast Fourier Transform
MLP	Multilayer perceptron;
MP3	MPEG-1/2 Audio Layer 3
PMC	Perceptron Multicamadas
RNA	Rede Neural Artificial
UFV	Universidade Federal de Viçosa
WAV	Waveform Audio File Format

Sumário

1	INTRODUÇÃO	12
1.1	Objetivo geral	14
1.2	Objetivos específicos	14
2	REFERENCIAL TEÓRICO	15
2.1	Forró e suas características	15
2.1.1	Principais vertentes do forró	15
2.1.2	Passo base da dança	15
2.1.3	Marcação da zabumba	16
2.2	Transformada de Fourier	17
2.3	Rede Perceptron Multicamadas	18
3	MATERIAIS E MÉTODOS	20
3.1	Materiais	20
3.2	Métodos	23
3.2.1	Banco de dados	24
3.2.2	Entradas para a rede PMC	24
3.2.3	Levantamento das saídas esperadas	25
3.2.4	Treinamento da rede PMC	25
4	RESULTADOS E DISCUSSÃO	28
4.1	Resultados com os arquivos em MP3	28
4.1.1	Treinamento	28
4.1.2	Desempenho da melhor rede	29
4.2	Resultados com os arquivos em WAV	29
4.2.1	Treinamento	29
4.2.2	Desempenho da melhor rede	30
4.3	Resultados com o banco misto, MP3 e WAV	30
4.3.1	Treinamento	30
4.3.2	Desempenho da melhor rede	32
4.4	Comparação das três abordagens	35
5	CONSIDERAÇÕES FINAIS	38
	REFERÊNCIAS	39

1 Introdução

No Brasil, entre 2006 e 2016, o Governo Federal intensificou os investimentos em educação, sobretudo no ensino público superior. Políticas como o Programa Universidade para Todos (PROUNI) (BRASIL, 2005), o Programa de Estímulo à Reestruturação e ao Fortalecimento das Instituições de Ensino Superior (Proies) (BRASIL, 2012a), o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI) (BRASIL, 2007), o Sistema de Seleção Unificada (SISU) (BRASIL, 2012b), o Fundo de Financiamento ao Estudante do Ensino Superior (FIES) (BRASIL, 2001), a Lei das Cotas nas Instituições Federais (BRASIL, 2012c), o Plano Nacional de Assistência Estudantil (PNAES) (BRASIL, 2010), a criação e expansão de universidades, a política de interiorização e a criação dos Institutos Federais de Educação, mudaram de forma acelerada o perfil socioeconômico dos estudantes do ensino superior (RISTOFF, 2014). Recentemente, entrou em vigência a Lei Nº 13.409 (2016) (BRASIL, 2016), que altera a “Lei das Cotas” (Lei Nº 12.711, 2012) (BRASIL, 2012c), a partir disso a tendência é que alunos surdos estejam mais presentes nas universidades públicas.

É natural pensar que o aumento da diversidade cultural dos estudantes universitários exija, das instituições, estudos que resultem em projetos compromissados em acolher as novas demandas dos novos grupos que constituem e constituirão o atual e futuro público universitário. Neste contexto, este projeto centra-se na inclusão do surdo na sociedade, iniciando pela comunidade universitária, entendendo que além de práticas que se preocupam com a aprendizagem dos conteúdos curriculares, a inclusão do surdo perpassa pela sua inserção na cultura local (LACERDA, 2006).

A dança tem forte presença em muitas culturas pelo mundo. No Brasil, o forró, uma festa que acabou sendo transformada em um estilo musical, é dançado por todas as camadas da sociedade e, principalmente, pela comunidade universitária. Existem vários projetos que promovem aulas de dança e eventos culturais, onde ocorrem a prática deste estilo de dança, integração entre estudantes e troca de conhecimentos. Apesar do reduzido número de surdos no ambiente universitário, e destes serem estigmatizados como seres não musicais (PAULA; PEDERIVA, 2017; HAGUIARA-CERVELLINI, 2003), alguns alunos surdos participam desses espaços. Na Universidade Federal de Viçosa (UFV), no projeto “Dança nas Moradias”, alunos do Curso de Dança da UFV oferecem aulas de dança de forró e samba de gafeira para os moradores dos alojamentos, onde alunos surdos participam de ambas modalidades oferecidas pelo projeto.

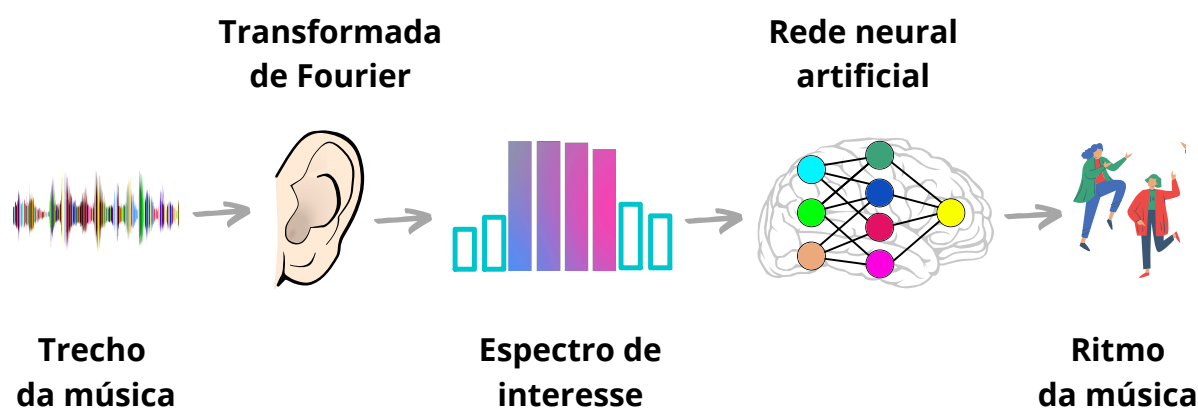
O aprendizado destes alunos consiste na visualização dos passos ensinados e o ritmo é percebido pela ajuda do parceiro e pela percepção das vibrações do som, quando estão

próximos das caixas de som. Trabalhos atuais mostram que é possível potencializar o contato do surdo com a música através de estímulos visuais e táteis (SHARP et al., 2019; BOSSEY, 2020; SHARP; BACON; CHAMPOUX, 2020; MIRZAEI; KÁN; KAUFMANN, 2020). Entretanto, para que a música possa ser convertida de uma onda sonora para luzes ou vibrações, não próximas aos alto-falantes, é necessário alguma ferramenta de interpretação, como a rede neural artificial (RNA).

A rede neural artificial é um algoritmo computacional treinado para classificar dados, aproximar funções e fazer previsões (SILVA; SPATTI; FLAUZINO, 2016). A literatura sobre processamento de áudio e reconhecimento de padrões utilizando aprendizado de máquinas é extensa, podendo ser citados trabalhos como Salamon e Bello (2017), Solanki e Pandey (2019), Cai e Cai (2019), Yu et al. (2020). Os trabalhos citados têm em comum a extração de características de áudio para fins de classificação através de técnicas como redes neurais artificiais. Porém, este trabalho diferencia-se por utilizar redes neurais para a estimação de um parâmetro da música, e não a classificação.

Neste trabalho, a rede neural foi treinada para estimar a duração do compasso musical. Este parâmetro poderá ser utilizado para sinalizar para pessoas surdas a velocidade da música reproduzida, a partir de um dispositivo. A RNA será responsável por extrair a marcação, necessitando de um dispositivo para receber o sinal e sinalizar marcação. A Figura 1 ilustra o modelo utilizado, com a transformada de Fourier para identificação das componentes espectrais da música e a rede neural como o sistema inteligente para identificar o ritmo da música através da duração do compasso. Vale ressaltar que este projeto se restringiu à elaboração do modelo, não foi o objetivo construir o dispositivo.

Figura 1 – Modelo implementado para a estimação do ritmo da música pelo compasso musical.



Fonte: Elaborado pelo autor.

Espera-se com este projeto facilitar a inserção do surdo na comunidade acadêmica,

através da dança (historicamente um importante agente inclusivo) (ZITOMER, 2017), almejando colaborar para inclusão efetiva do aluno. Consequentemente, a inserção deve promover a melhoria da sua qualidade de vida e seu permanecimento no ambiente universitário até o cumprimento de seus objetivos pessoais, seja na graduação ou na pós-graduação.

1.1 Objetivo geral

Este trabalho tem como objetivo propor um método para estimar a duração de um compasso musical de músicas de forró, com baixo custo computacional para fornecer a velocidade que uma pessoa surda deve dançar, a partir de estímulos sensoriais alternativos à audição.

1.2 Objetivos específicos

Os objetivos específicos do projeto são:

- Selecionar um banco de músicas de forró variado com as músicas mais tocadas no meio universitário;
- Levantar um banco de valores de duração do compasso esperado para as músicas selecionadas;
- Definir as características musicais para alimentar as entradas da rede neural;
- Selecionar a melhor RNA para estimar a duração do compasso das músicas selecionadas.

2 Referencial Teórico

Este capítulo contém uma teorização a respeito da escolha das músicas de forró que compõem o banco de dados e a fundamentação teórica dos métodos adotados no modelo de estimação de duração do compasso de músicas de forró.

2.1 Forró e suas características

A acepção do termo “forró” é múltipla, segundo [Junior e Volp \(2005\)](#), não existe um consenso sobre significado do termo e que provavelmente este consenso nunca ocorra. Em diálogos referindo-se ao forró, a palavra pode significar uma festa ou local e um gênero musical ([GIFFONI, 1997](#)). Neste trabalho são consideradas três componentes do fenômeno forró: as músicas que são tocadas, a forma de dançar e a festa que reúne as pessoas para ouvir e dançar xote, baião, xaxado e rasta pé ([NUNES; SOUZA, 2014](#)).

2.1.1 Principais vertentes do forró

O forró, enquanto gênero musical, pode ser classificado em três categorias: Forró Pé-de-serra, Forró Universitário e Forró Eletrônico ([JUNIOR; VOLP, 2005](#)). O Forró Pé-de-serra tem forte inspiração artística no universo rural do sertanejo e originou-se em meados da década 1940, no Nordeste. O Forró Universitário surgiu a partir do Forró Pé-de-serra com fortes influências do Rock’n Roll, do Samba, do Funk e do Reggae, introduzidas quando jovens sulistas começaram a tocar e cantar forró nas décadas de 1990/2000. Já o Forró Eletrônico tem como principal característica a inserção de instrumentos eletrônicos, mostrando uma linguagem estilizada e um visual chamativo, também originado na década 1990, esta vertente é pouco comum no contexto universitário.

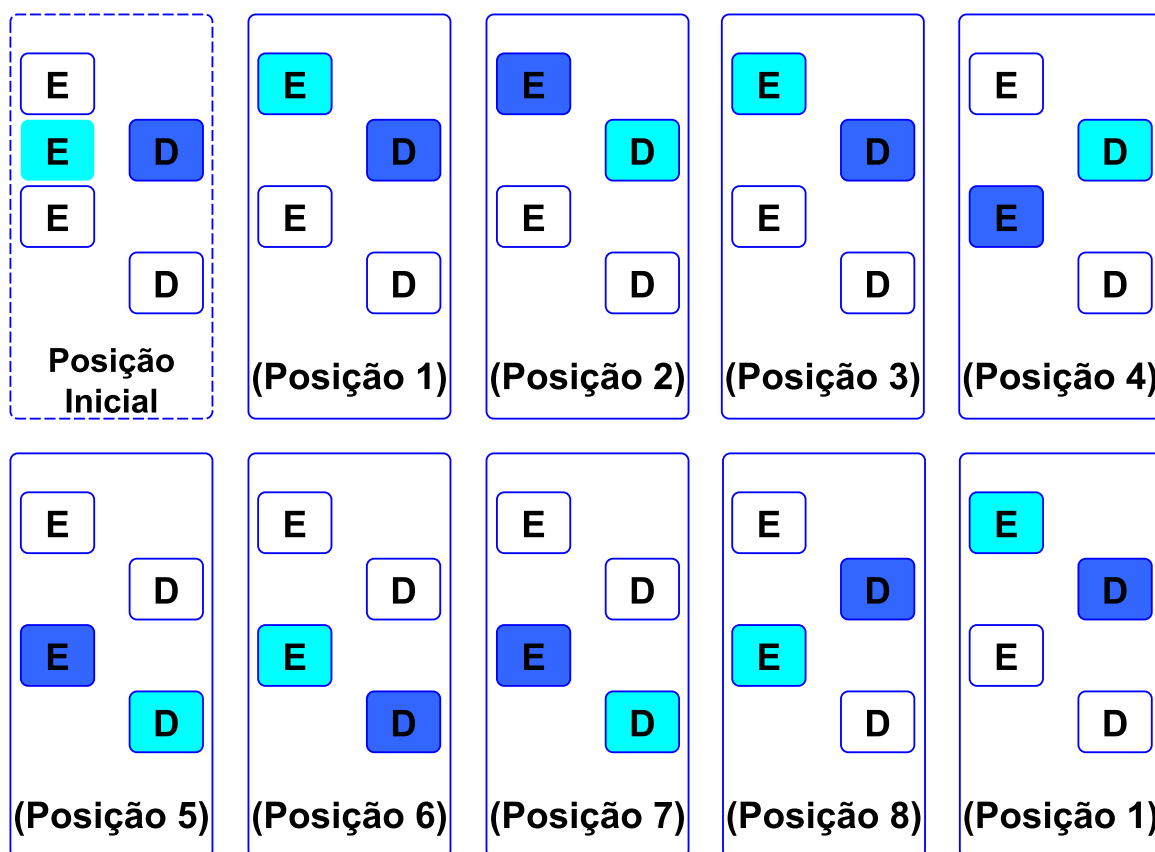
2.1.2 Passo base da dança

Segundo [Schoenberg \(1990\)](#), uma das formas de caracterizar o passo base da dança é em oito tempos (1, 2, 3, ..., 8). No forró, dois passos base são mais comuns o Passo Base Lateral e o Passo Base Frente e Trás, ambos são realizados ao longo de dois compassos ([JUNIOR et al., 2009](#)). A Figura 2 apresenta as posições dos pés na execução do Passo Base Frente e Trás e a transferência de peso entre os dois pés, conforme descrito por [Junior et al. \(2009\)](#). Os quadrados indicam a localização dos pés, considerando que a pessoa está dançando ocupando sempre o mesmo espaço. Para cada posição, o quadrado preenchido em azul escuro representa a posição do pé de sustentação, que está recebendo o peso do corpo, enquanto que o quadrado em azul claro representa a posição do pé que está

apenas tocando o chão. A sequência apresentada na Figura 2, caracteriza o movimento do condutor, este inicia com os pés paralelos (Posição inicial) e avança com o pé esquerdo a frente (Posição 1). Por outro lado, o conduzido inicia com os pés paralelos e recua o pé com o pé direito para trás (Posição 5).

O Passo Base Frente e Trás é executado em oito movimentos, deslocando de uma posição para outra (1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8 e 8-1), a movimentação foi pensada a partir do deslocamento do condutor, existe coincidência das Posições 1 e 8, refletindo a característica periódica do passo base. De forma análoga, o deslocamento do conduzido segue a sequência completa e a periodicidade, partindo da Posição 5 (5, 6, 7, 8, 1, 2, 3, 4, 5, ...).

Figura 2 – Diagrama do Passo Base Frente e Trás. O pé esquerdo (E) e direito (D) são representados pelos quadrados com as letras e as cores azul escuro e azul claro representam o pé de sustentação e o pé que toca o chão, respectivamente.



Fonte: (JUNIOR et al., 2009). Adaptado.

2.1.3 Marcação da zabumba

O forró, em geral, possui instrumentos e cantores. A faixa de frequência da voz humana varia entre 400 Hz e 10 kHz, tendo a maior quantidade de energia nas frequências próximas a 2 kHz (TITZE; MARTIN, 1998). A zabumba junto com a sanfona e o triângulo

formam a estrutura básica instrumental do Forró Pé de Serra e do Forró Universitário (JUNIOR; VOLP, 2005; NUNES; SOUZA, 2014). Os instrumentos de percussão vibram em baixas frequências, na faixa de 70 Hz a 1 kHz, tendo a maior densidade de energia dependendo do instrumento. A zabumba possui maior densidade de energia nas frequências entre 50 Hz e 300 Hz e é utilizada como referência de marcação da música, indicando a velocidade de execução dos passos do forró.

2.2 Transformada de Fourier

Toda onda pode ser caracterizada pela frequência na qual oscila. No caso das ondas sonoras, elas podem ser caracterizadas por várias frequências ao mesmo tempo, cada uma com uma quantidade de energia específica. No ouvido interno, o som é recebido pela cóclea que é um analisador de frequência hidromecânico, cujo principal papel é realizar uma decomposição espectral em tempo real do sinal acústico recebido (WHITE; GROSH, 2005). A transformada de Fourier mede a energia em todo o espectro de frequência da onda. Para um sinal contínuo $x(t)$, sua transformada pode ser definida por

$$X(j\omega) = \int_{-\infty}^{\infty} s(t)e^{-j\omega t} dt \quad (2.1)$$

e para um sinal discreto $x[n]$, sua transformada pode ser definida por

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x[n]e^{-j\omega n} \quad (2.2)$$

(OPPENHEIM; WILLISKY, 2010).

A transformada de Fourier também pode ser representada pelos seus componentes real e imaginário, ou em termos de magnitude e fase. A magnitude-fase para a transformada de Fourier de tempo discreto em (2.2), é representada por

$$X(e^{j\omega}) = |X(e^{j\omega})| e^{j\angle X(e^{j\omega})} \quad (2.3)$$

em termos do módulo e da fase (OPPENHEIM; WILLISKY, 2010).

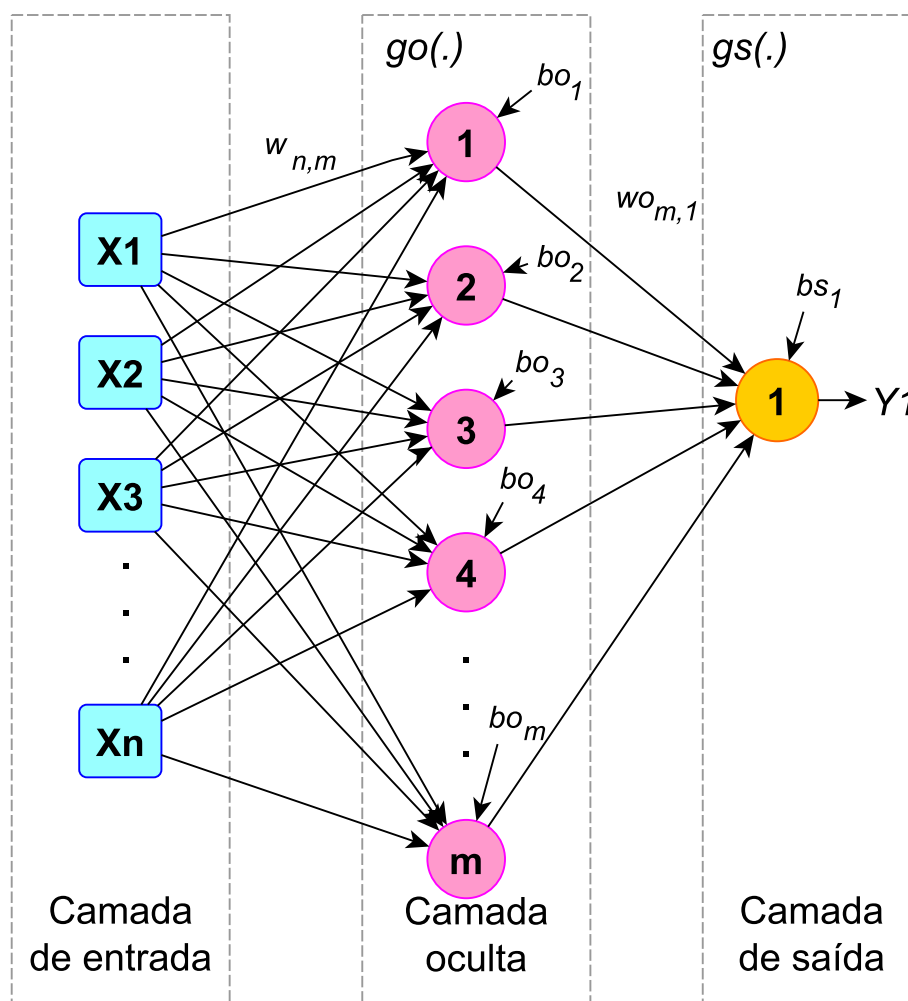
O módulo ($|X(e^{j\omega})|$) representa a magnitude de energia contida em cada frequência do sinal, enquanto que a fase ($\angle X(e^{j\omega})$) está relacionada à natureza do sinal (OPPENHEIM; WILLISKY, 2010).

Computacionalmente, a Transformada Rápida de Fourier (FFT – Fast Fourier Transform) é amplamente utilizada por ser um algoritmo matemático eficiente para a obtenção das amplitudes e fases dos diversos componentes de um sinal discreto (COOLEY; TUKEY, 1965).

2.3 Rede Perceptron Multicamadas

As redes neurais partem da modelagem matemática do neurônio biológico (MC-CULLOCH; PITTS, 1990). Uma rede neural é um algoritmo que interpreta informações de sinais de entrada e entrega uma saída significativa ao usuário. A Figura 3 apresenta a estrutura de uma rede neural com uma camada oculta.

Figura 3 – Estrutura de uma rede PCM com uma camada oculta.



Fonte: Elaborado pelo autor.

As entradas da rede neural são as variáveis $\mathbf{x} = [x_1 \ x_2 \ x_3 \dots x_n]^T$ e a saída é o Y_1 . Os termos $bo_1, bo_2, bo_3, \dots, bo_m$ e bs_1 são denominados potenciais de ativação. Os termos $w_{n,m}$ e $w_{s,m,1}$ são denominados pesos sinápticos. Os potenciais de ativação representam o valor que a soma dos sinais de entrada, ponderados pelos pesos sinápticos de cada entrada, deve atingir em cada neurônio para que aquela informação seja repassada para a próxima camada da rede. Finalmente, os termos $go(.)$ e $gs(.)$, representam as funções de ativação para a camada oculta e de saída respectivamente. Haykin (2007), destaca três funções

de ativação comuns na implementação de redes neurais a função logística, a tangente hiperbólica e a linear.

A rede neural precisa ser treinada para que possa interpretar as informações recebidas na camada de entrada. O processo de treinamento supervisionado da rede neural consiste em apresentar um conjunto de entradas combinados com saídas desejadas. A saída desejada é o valor que a rede neural deve retornar ao receber o conjunto correspondente de entrada. Cada um dos m neurônios da camada oculta possui pesos que ponderam a informação apresentada por uma das n entradas. Estes coeficientes são ajustados no processo de treinamento até que a rede neural apresente resultados satisfatórios de generalização, podendo ser utilizada, em um segundo momento, na fase de operação (SILVA; SPATTI; FLAUZINO, 2016).

A estimação de um parâmetro utilizando treinamento supervisionado consiste em apresentar para a rede um par de entradas e saídas, de modo que o treinamento possa reproduzir essa relação de entrada e saída (SILVA; SPATTI; FLAUZINO, 2016). O treinamento supervisionado da rede PCM utilizando *backpropagation* consiste em duas etapas. Na primeira, *forward* a rede é alimentada na camada de entrada e, a partir desta camada as unidades calculam sua resposta que é produzida na camada de saída e o erro é calculado. Na segunda etapa, *backward*, o erro é propagado da camada de saída para a camada de entrada, e os pesos das conexões das unidades das camadas internas vão sendo atualizados segundo o erro quadrático médio, também conhecido como regra delta (HAYKIN, 2007). Esse processo de atualização de pesos requer um algoritmo otimizado, de forma a garantir a convergência da rede com o baixo custo computacional. O Levenberg-Marquardt é um algoritmo de retro-propagação eficiente e confiável para a convergência do treinamento (MORE, 1978).

Depois da fase de treinamento, a rede PMC é validada, com amostras que não foram apresentadas durante o processo de treinamento. Se o erro durante a fase de validação não foi satisfatório, os pesos sinápticos e os limiares de ativação são ajustados e uma nova fase de treinamento é iniciada. Caso a rede tenha uma boa capacidade de generalização, estimando corretamente as saídas para o conjunto de dados na fase de validação, a rede passa para uma fase de teste (SILVA; SPATTI; FLAUZINO, 2016). A fase de teste serve para avaliar o real desempenho da rede diante entradas desconhecidas durante o treinamento e a validação. É na fase de teste que a rede mostra o real potencial para estimação da saída, conforme o objetivo do treinamento.

3 Materiais e Métodos

Neste capítulo são apresentados os materiais utilizados e os métodos adotados para modelar a estimação da duração do compasso musical.

3.1 Materiais

Para este trabalho foram selecionadas 82 músicas em MP3 (MPEG-1/2 Audio Layer 3) com taxa de amostragem 44.100 Hz das vertentes Forró Pé-de-serra e Forró Universitário para compor o banco de dados (Tabela 1). Para garantir a generalização dos resultados, foram selecionadas músicas com andamentos distintos, de músicas “mais lentas” à músicas “mais rápidas”. É importante ressaltar que a seleção das músicas não almejou uma separação precisa entre Forró Pé-de-serra e Universitário, por não ser necessidade do trabalho. A escolha das músicas se deu pela variedade rítmica e a popularidade dos gêneros no contexto universitário, sendo orientada por uma instrutora de dança.

Figura 4 – Espaço de dança “Forró do Itaú” no dia da gravação das músicas.



Fonte: Foto tirada pelo autor.

Das 82 músicas selecionadas, 37 músicas, estão marcadas com “*” Tabela 1, foram reproduzidas no espaço de dança conhecido como “Forró do Itaú” (4), realizado todos os domingos no Campus Viçosa da UFV pelo projeto “Forró Pé de Serra no Campus”. As músicas foram gravadas por celular em formato WAV (Waveform Audio File Format), com taxa de amostragem 44.100 Hz. Como a gravação teve o intuito de simular uma situação

cotidiana, as músicas foram gravadas com uma pessoa dançando com o celular no bolso, no dia 08/12/2019.

A seleção das músicas apresentada na Tabela 1 inclui seus nomes, os interpretes das versões utilizadas e a respectiva duração do compasso. As músicas que também foram regravadas no “Forró do Itaú” são encontradas em destaque “*”. A extração das características das músicas, o treinamento da PMC e a avaliação dos resultados obtidos foram realizados no *software* Matlab 2016a.

Tabela 1 – Músicas do banco de dados com o interprete da versão utilizada e a duração do compasso.

Música	Intérprete	Compasso (s)
De Mala e Cuia	Camarão	1,002
O Casamento da Raposa	Nicolas Krassik	1,033
É Proibido Cochilar*	Circuladô de Fulô	1,090
Menina Bela	Trio Maracaia	1,108
Baião Caçula*	Dominguinhos	1,121
Forró Animado	Marinalva	1,131
De Tamanco Mulher*	Trio Forrozão	1,139
Enxuga o Rato	Zé Ranulfo de S. Talhada	1,140
Não Tô Nem Ai	Os Filhos de Vitor	1,160
Baião Jovem*	Trio Dona Zefa	1,164
Plantio de Amor	Dominguinhos	1,165
Novinho Em Folha	Trio Xamego	1,179
Seu Siriri	Trio Dona Zefa	1,189
Evolução	Ary Lobo	1,193
Canto do Sabiá*	Trio Dona Zefa	1,204
Tem Problema Não*	Trio Bastião	1,206
O Melhor Forró do Munundo	Os 3 do Nordeste	1,209
O Baião Que Deu	David Cruz	1,216
Forró do Bole Bole	Trio Nordestino	1,224
Você vai se Machucar*	Trio Dona Zefa	1,238
Tem Quem Reze Por Mim	Trio Juazeiro	1,238
Avisa*	Falamansa	1,243
Agradar Você	Trio Potiguá	1,244
Alambique de Barro*	Os 3 do Nordeste	1,245
Cigano Terrorista*	Trio Cangaço	1,246
Cantiga do Sapo	Jackson do Pandeiro	1,247
Tome Forró	Dona Zaíra	1,248
Esconderijo do Amor*	Genaro	1,249

*Músicas que foram reproduzidas e regravadas no “Forró do Itaú”.

Música	Intérprete	Compasso (s)
Coração Molim	Luiz Gonzaga	1,250
Eu Balanço	Jackson do Pandeiro	1,251
Sete meninas	Jackson do Pandeiro	1,252
É Bom Pra Valer	Os 3 do Nordeste	1,259
Sereia do Mar	Camarão	1,261
Matilde	Forrócana	1,263
Ta Esquentando*	Trio Dona Zefa	1,264
Pé do Lajeiro	João do Vale	1,264
Briga de Artista*	Trio Dona Zefa	1,264
Eu sou assim	Marinalva	1,267
O Sertão Te Espera	Dominguinhos	1,272
Só Via Faca Voar	Bezerra da Silva	1,275
Tenente Bezerra	Genival Lacerda	1,289
Baioneiro Gonzagão*	Trio Dona Zefa	1,290
Chuva Miúda	Ary Lobo	1,298
Sambarcando	João do Pife	1,299
Samba da Boca do Mato	Eliano Julião	1,310
Me Chamo Forró	Ó do Forró	1,320
Até o Dia Clarear	Os 3 do Nordeste	1,323
Baião de Respeito*	Dona Zaira	1,333
Xote dos Milagres*	Falamansa	1,347
Começaram a Gritar	Ó do Forró	1,349
Desilusão*	Dominguinhos	1,363
Gabriela	Trio Meketréfe	1,365
Minha Dita é Desse Jeito	Cachimbinho	1,366
Produto Nacional	Trio Dona Zefa	1,367
Asas*	Bicho de Pé	1,375
Amor de Mentirinha	Jackson do Pandeiro	1,377
Vida Boa Danada*	Trio Dona Zefa	1,381
Dona Mariquinha	A Barca	1,391
Menino Prodígio	Ary Lobo	1,396
Anjo da Guarda Noturno*	Bicho de Pé	1,424
Amanhecer*	Os 3 Cangaceiros	1,449
Baião*	Balança Eu	1,453
Súplica Cearense	Gordurinha	1,454
100 anos*	Falamansa	1,454

*Músicas que foram reproduzidas e regravadas no “Forró do Itau”.

Música	Intérprete	Compasso (s)
De Mala e Cuia	Trio Virgulino	1,463
Encosta n'Eu*	Estakazero	1,501
Bata a Porta*	Black	1,543
Lilith*	Alcalyno	1,570
Medo de Escuro*	Falamansa	1,582
Vem me Ver*	Circuladô de Fulô	1,582
Amor de Rede	Trio Jurití	1,600
Xote Universitario*	Falamansa	1,601
Fino Trato	Forró na Contramão	1,645
Você Já Fez Morada Em Meu Juízo	Trio Xamego	1,646
Nosso xote*	Bicho de Pé	1,666
Acalanto*	Black e Bárbara Greco	1,666
Eu e Você*	Bicho de Pé	1,675
Como Todo Amor*	Forrueiros	1,714
Nossos Olhos*	Forró Massapê	1,716
Casinha Pronta*	Black	1,752
Feitiço da Lua*	Forrueiros	1,795
Flor de Lis*	Circuladô de Fulô	1,892

Fonte: Elaborado pelo autor.

3.2 Métodos

Este trabalho almejou estimar a duração do compasso de uma música de forró a partir do espectro de frequência de uma amostra três segundos da música a partir de uma rede perceptron multicamadas. A modelagem consistiu em criar os bancos de dados a partir das músicas em MP3 e WAV (gravadas), processamento do banco de dados para fornecer as entradas para a rede neural, levantamento da duração do compasso esperado para cada música e o treinamento da rede neural. As etapas do modelo proposto são detalhadas a seguir.

3.2.1 Banco de dados

Foram adotados três bancos de dados para a estimação da duração do compasso das músicas:

- 82 arquivos de áudio em MP3;
- 37 arquivos de áudio em WAV (músicas gravadas);
- 119 arquivos de áudio 82 em MP3 e 37 em WAV;

O terceiro banco de dados foi adotado sob a hipótese de que oferecer à rede as mesmas músicas com e sem ruído poderia capacitá-la para distinguir as entradas fundamentais das entradas ruidosas.

A preparação das músicas para o banco de dados consistiu em retirar o início e o final das músicas e normalizar a amplitude do sinal em todo domínio do tempo. A normalização foi feita dividindo-se cada ponto pelo valor eficaz da amostra V_{rms} , definido por (3.1), para que eventual disparidade entre o volume das músicas não interfira nos resultados. Em seguida, cada música foi segmentada em trechos de três segundos com sobreposição de um segundo entre os segmentos.

$$V_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N V_i^2} \quad (3.1)$$

A cada segmento foi atribuído uma duração do compasso da música a qual pertencia. Para isso, foi considerado que a duração do compasso das músicas escolhidas não varia ao longo da música. Essa consideração, simplifica o método por permitir adotar a média dos tempos cronometrados para cada amostra de uma mesma música. Entretanto, a mesma consideração pode diminuir a capacidade de estimação da rede neural.

3.2.2 Entradas para a rede PMC

As entradas para a rede neural foram as principais componentes do espectro de frequência da zabumba (50 a 300 Hz) retiradas a partir do espectro completo de cada trecho segmentado da música. O espectro foi dividido em 25 faixas com espaçamento de 10 Hz. Assim, cada entrada corresponde à médias das amplitudes das componentes nas frequências correspondente aquela faixa. Quanto maior a amplitude da componente na frequência, maior a ocorrência do som caracterizado por esta frequência. É esperado que músicas mais rápidas, nas quais a zabumba é tocada em uma velocidade maior, as amplitudes das componentes na faixa da zabumba sejam maiores do que em músicas mais lentas e que este efeito seja suficiente para que a rede consiga estimar a duração do

compasso. O uso das componentes espectrais é comum para reconhecimento e classificação de sinais de áudio (BHALKE; RAO; BORMANE, 2016; YU et al., 2020).

3.2.3 Levantamento das saídas esperadas

A extração do tempo do compasso foi feita de forma indireta a partir do tempo gasto para a execução de um passo base de forró. Para medir o tempo do compasso de cada música selecionada, uma instrutora de forró dançou todas as músicas seguindo o Passo Base Frente e Trás. A medição consistiu em cronometrar o tempo gasto para realizar o movimento, utilizando a Posição 8 (veja Figura 2) como referência, ou seja, toda vez que a instrutora pisava com o pé direito a frente, o cronômetro era reiniciado manualmente. Foram feitas 20 medições para cada música, adotando-se a média como sendo o tempo de execução do passo base da música observada. Finalmente, o tempo de duração do compasso se deu pela divisão do tempo de duração do passo base por dois. A duração do compasso medida para cada música pode ser observada na Tabela 1.

3.2.4 Treinamento da rede PMC

A topologia de rede escolhida foi o perceptron multicamadas (PMC) com camada única, devido sua característica de aproximação universal de funções (SILVA; SPATTI; FLAUZINO, 2016). Outra característica desta rede é mapear qualquer função contínua no espaço das funções reais, desde que, seja utilizada uma função de ativação contínua e limitada em sua imagem (SILVA; SPATTI; FLAUZINO, 2016). Os principais parâmetros da arquitetura e treinamento da rede PMC são:

- Número de entradas: 25;
- Número de saídas: 1;
- Função de ativação da camada oculta: Tangente hiperbólica;
- Função de ativação da camada de saída: Linear;
- Algoritmo de treinamento: Levenberg-Marquardt;
- Repartição do banco de dados: 70% treinamento, 15% validação e 15% teste.

A partir dos parâmetros estabelecidos, para encontrar a rede que melhor se adaptou ao problema foram treinadas 40 redes, variando somente o número de neurônios na camada oculta de 11 a 51 conforme o método de Fletcher-Gloss descrito por

$$2\sqrt{n} + n_2 \leq n_1 \leq 2n + 1 \quad (3.2)$$

onde n é o número de entradas, n_1 é a quantidade de neurônios na camada oculta e n_2 é o número de neurônios na camada de saída (SILVA; SPATTI; FLAUZINO, 2016).

Cada rede, ou seja determinada quantidade de neurônios na camada oculta, foi treinada 100 vezes com novos sorteios dos pesos, potenciais de ativação e sorteio das amostras para treinamento, validação e teste. Para evitar *overfitting*, o treinamento de cada rede foi finalizado sempre que o Erro Quadrático Médio (EQM) dos dados de validação paravam de diminuir. O EQM pode ser definido por

$$\text{EQM} = \frac{1}{N} \sum_{i=1}^N (Yd_i - Y_i)^2, \quad (3.3)$$

na qual N é o número de amostras durante o treinamento, Yd_i é a saída desejada e Y_i é a saída encontrada pela rede rede PMC. Foi considerada a melhor rede aquela que obteve o menor Erro Percentual Médio (EPM) das 40 redes avaliadas treinamentos realizados. O EPM pode ser definido por

$$\text{EPM}(\%) = \frac{1}{N} \sum_{i=1}^N \frac{|Yd_i - Y_i|}{Yd_i} \times 100\%, \quad (3.4)$$

com Yd_i o valor esperado, Y_i o valor encontrado e N o número de amostras.

Para medir a dispersão do EPM encontrado pelas redes treinadas foi utilizado o Desvio Padrão (DP) descrito por

$$\text{DP}(\%) = \sqrt{\sum \frac{(\text{EP}_i - \bar{\text{EP}})^2}{N}} \quad (3.5)$$

com $\bar{\text{EP}} = \text{EPM}$ a média do Erro Percentual (EP), EP o valor do erro percentual para cada amostra e N o número de amostras.

Para encontrar a melhor configuração (números de neurônios na camada oculta) de rede capaz de estimar a duração do compasso do forró, as redes treinadas a partir de cada banco de dados foram submetidas à três etapas de teste. A primeira consistiu em avaliar a melhor configuração de rede entre as treinadas com o mesmo tipo de banco de dados, a partir desta etapa a melhor arquitetura de rede de cada um dos três treinamentos passaram por mais duas etapas de teste que consistiu em comparar as três melhores redes com um mesmo conjunto de amostras do banco em MP3 e WAV.

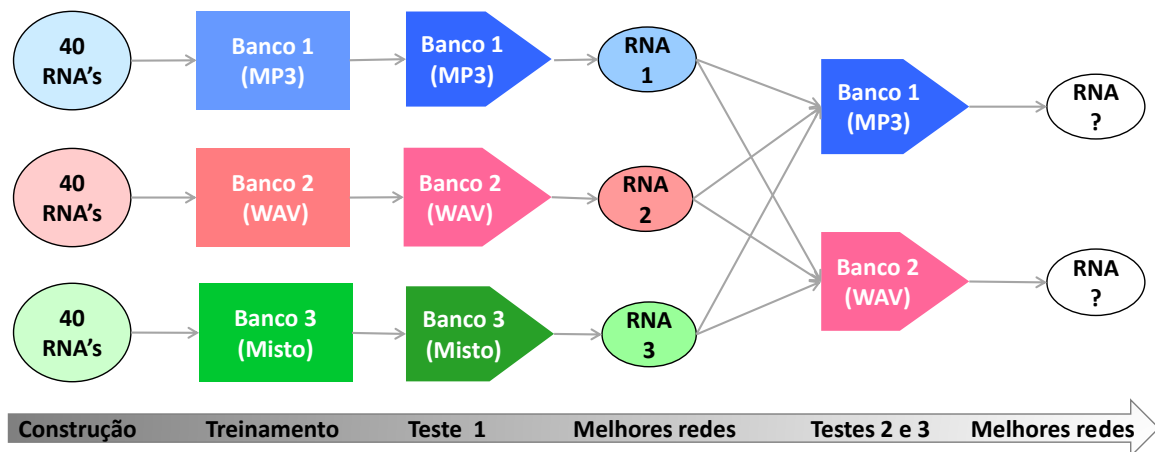
Conforme ilustrado na Figura 5 foram utilizados três bancos de dados para as etapas de teste. O Banco de Teste 1 foi utilizado nos Testes 1 e 2 enquanto que o Banco do Teste 2 foi utilizado para os Testes 1 e 3. Desta forma, foi possível comparar o desempenho das Redes 1, 2 e 3 para os dois tipos de bancos de dados sem a utilização de novos arquivos de músicas. Testes listado a seguir:

- **Teste 1:** Neste teste, das 120 redes foram escolhidas 3 redes candidatas, que foram as melhores redes para cada um dos três bancos de dados. No Teste 1 as redes foram

submetidas à amostras da mesma natureza das amostras do treinamento, ou seja as redes treinadas com o Banco 1 de treinamento, foram testadas com o Banco 1 de teste.

- **Teste 2:** As três redes selecionadas no Teste 1 tiveram seus desempenhos comparado quando submetidas ao Banco 1 de teste, músicas em MP3 (Figura 5). No final, a rede escolhida no Teste 2 será a mais apta para estimar a duração do compasso a partir de músicas em MP3.
- **Teste 3:** As três redes selecionadas no Teste 1 tiveram seus desempenhos comparado quando submetidas ao Banco 2 de teste, músicas em WAV (Figura 5). No final, a rede escolhida no Teste 3 será a mais apta para estimar a duração do compasso a partir de músicas em WAV (com ruídos).

Figura 5 – Processo de seleção das melhores redes para a estimação do compasso de músicas de forró para as músicas em MP3 e WAV.



Fonte: Elaborado pelo autor.

4 Resultados e Discussão

Neste capítulo é apresentado os resultados obtidos na etapa de teste para cada um dos três treinamentos (Teste 1) feitos com músicas em MP3, WAV e com o banco de dados misto. Ao final é apresentado a comparação dos resultados obtidos com as melhores redes de cada treinamento quando submetidas ao mesmo conjunto de amostras de teste de músicas em MP3 (Teste 2) e gravadas em WAV (Teste 3).

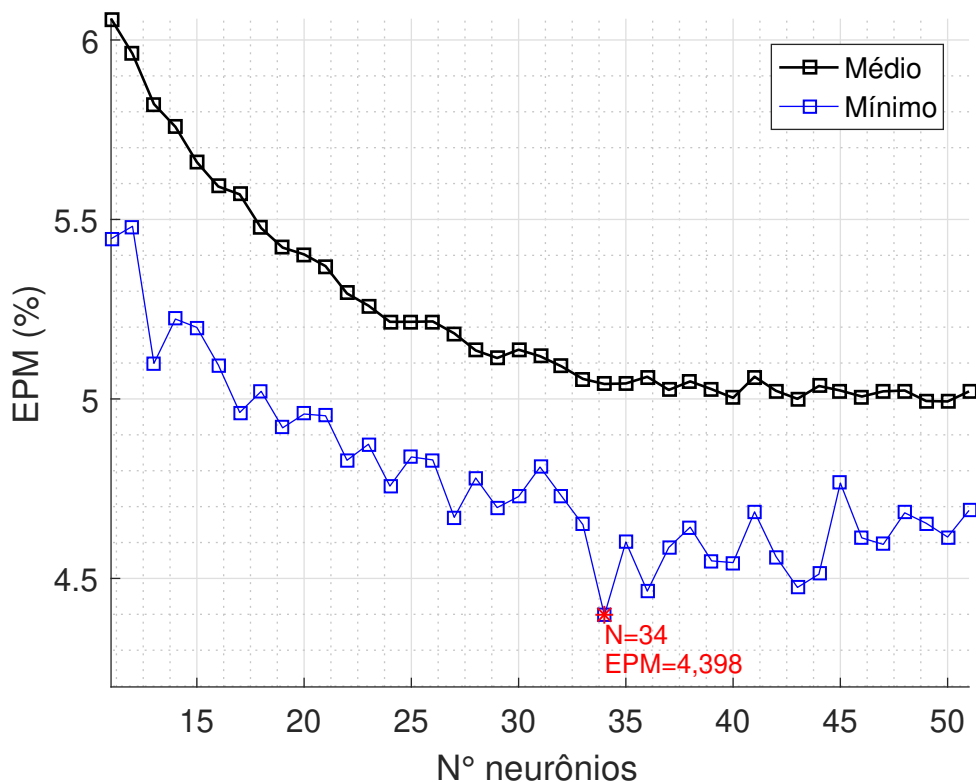
4.1 Resultados com os arquivos em MP3

As 82 músicas em formato MP3 quando segmentadas totalizaram 6433 amostras. As durações dos compassos variaram de 1,002 s até 1,892 s.

4.1.1 Treinamento

Na Figura 6 pode-se observar que a rede PMC com 34 neurônios na camada oculta foi a que obteve melhor desempenho para estimar a duração do compasso das músicas em MP3, com $EPM = 4,398\%$, para a fase de teste.

Figura 6 – EPM das redes PMC, para as músicas em MP3, em função do número de neurônios na camada oculta.

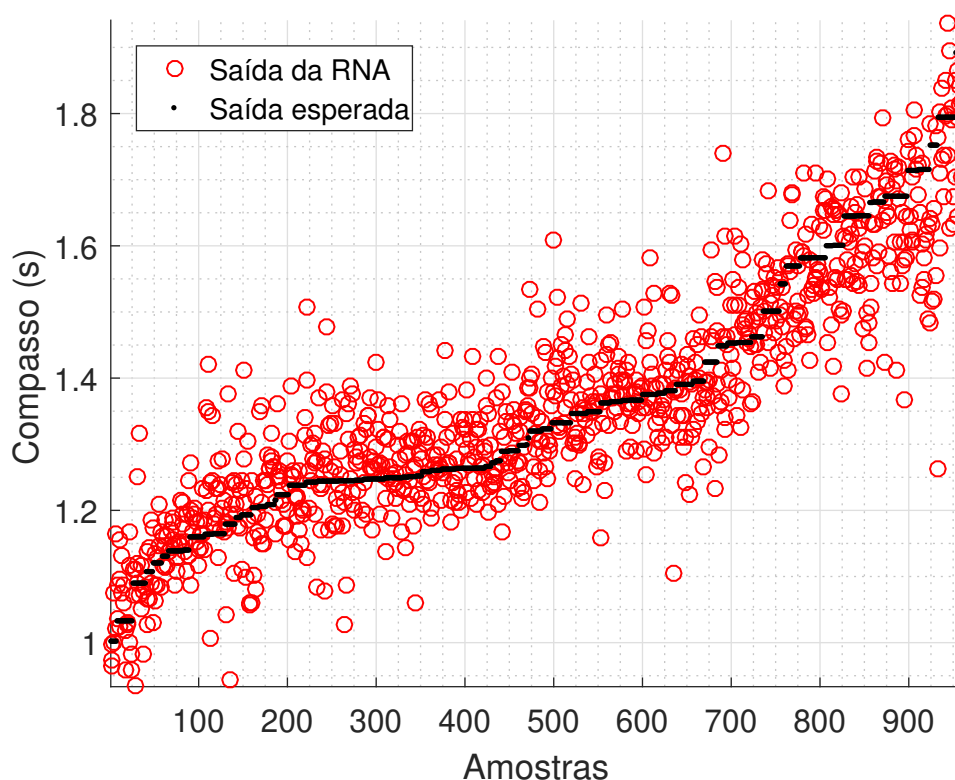


Fonte: Elaborado pelo autor no Matlab.

4.1.2 Desempenho da melhor rede

Na Figura 7, é apresentado a resposta obtida pela Rede 1 (34 neurônios na camada oculta destacada na Figura 6) em comparação com os valores esperados, para as amostras de teste. É possível perceber aproximação entre os valores obtidos e os esperados. Na Figura 8 é apresentado a capacidade de aproximação da Rede 1, na qual o erro percentual médio das amostras foi $EPM = 4,398 \pm 3,887 \%$.

Figura 7 – Saída estimada para os dados de teste pela rede com o menor EPM (Rede 1) .



Fonte: Elaborado pelo autor no Matlab.

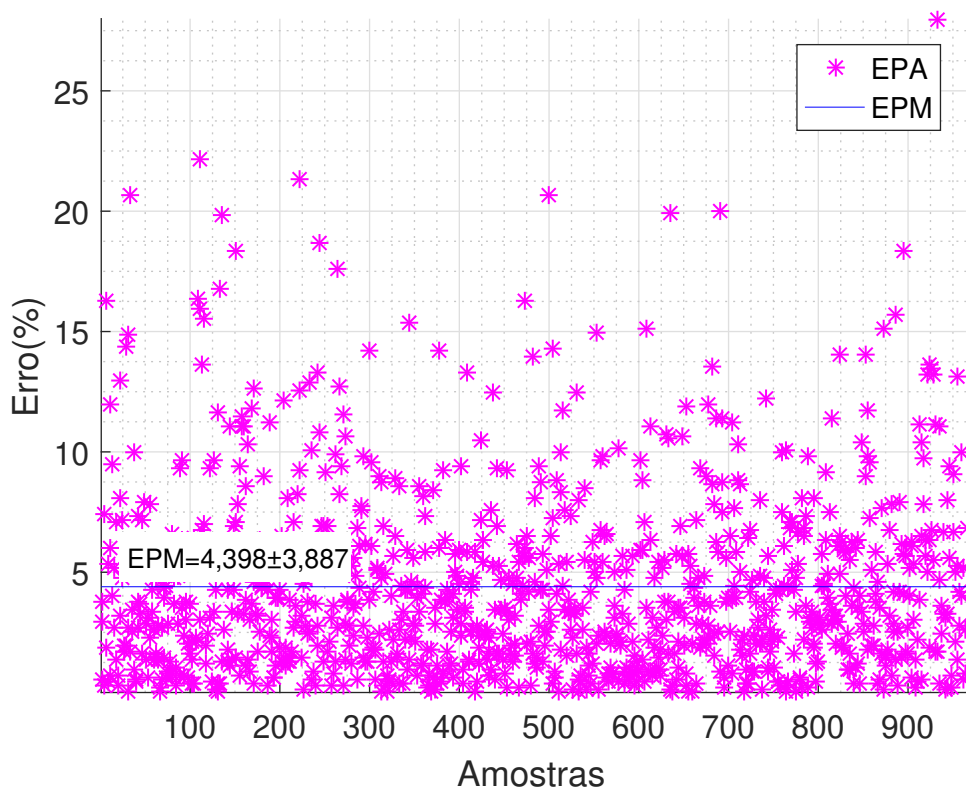
4.2 Resultados com os arquivos em WAV

As 37 músicas em formato WAV quando segmentadas totalizaram 3200 amostras. As durações dos compassos variaram de 1,090 s até 1,892 s.

4.2.1 Treinamento

Na Figura 9 pode-se observar que rede PMC com 31 neurônios na camada oculta foi a que obteve melhor desempenho para estimar a duração do compasso das músicas em WAV com $EPM = 5,293\%$, para a fase de teste.

Figura 8 – Erro obtido com os dados de teste pela rede com o menor EPM (Rede 1), referente à Figura 7.



Fonte: Elaborado pelo autor no Matlab.

4.2.2 Desempenho da melhor rede

Na Figura 10, é apresentado a resposta obtida pela Rede 2 (31 neurônios na camada oculta destacada na Figura 9) em comparação com os valores esperados, para as amostras de teste. É possível perceber aproximação entre os valores obtidos e os esperados. Na Figura 11 é apresentado a capacidade de aproximação da rede PMC, na qual o erro percentual médio das amostras foi $EPM = 5,293 \pm 4,215 \%$.

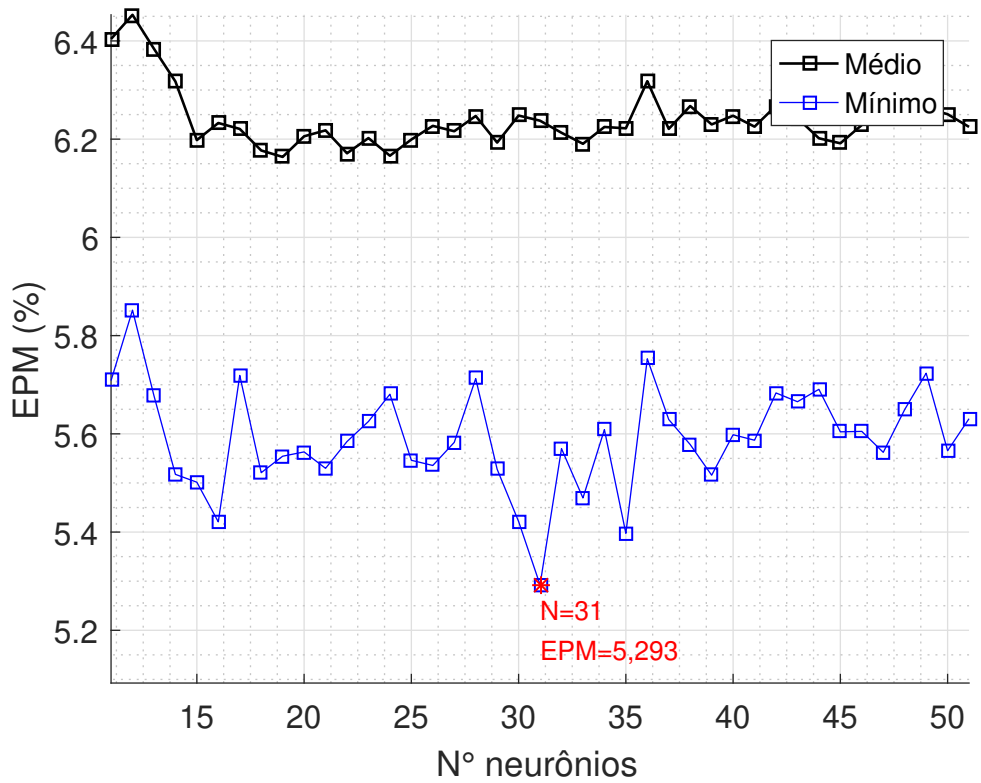
4.3 Resultados com o banco misto, MP3 e WAV

O conjunto misto contou com 37 músicas em WAV e 82 músicas em MP3 quando segmentadas as músicas totalizaram 9633 amostras. As durações dos compassos variaram de 1,002 s até 1,892 s.

4.3.1 Treinamento

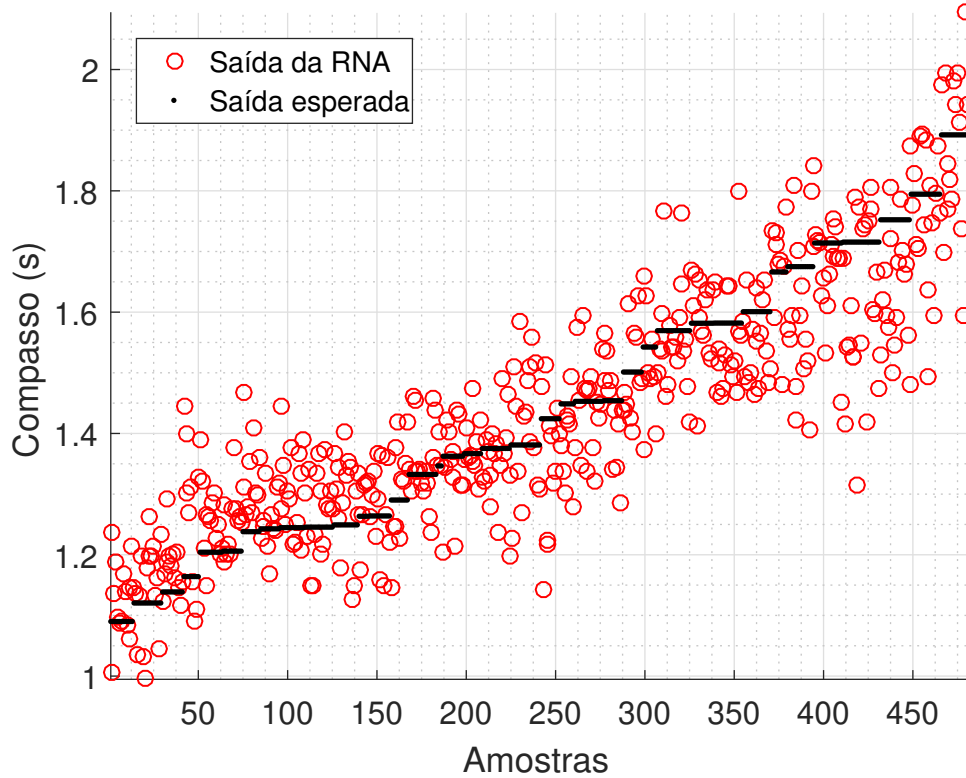
Na Figura 9 pode-se observar que a rede PMC com 38 neurônios na camada oculta foi a que obteve melhor desempenho para estimar a duração do compasso das músicas do banco misto com $EPM = 5,243\%$, para a fase de teste.

Figura 9 – EPM das redes PMC, para as músicas em WAV, em função do número de neurônios na camada oculta.



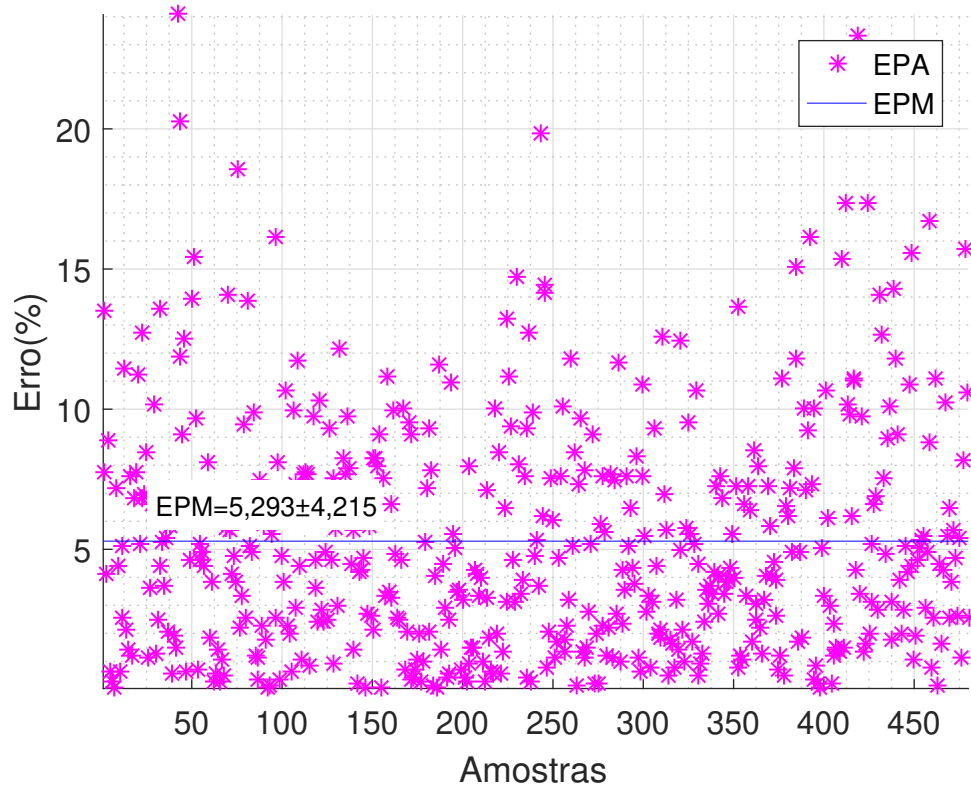
Fonte: Elaborado pelo autor no Matlab.

Figura 10 – Saída estimada para os dados de teste pela rede com o menor EPM (Rede 2).



Fonte: Elaborado pelo autor no Matlab.

Figura 11 – Erro obtido com os dados de teste pela rede com o menor EPM (Rede 2), referente à Figura 10.



Fonte: Elaborado pelo autor no Matlab.

4.3.2 Desempenho da melhor rede

Na Figura 13, é apresentado a resposta obtida pela da Rede 3 (38 neurônios na camada oculta destacada na Figura 12) em comparação com os valores esperados, para as amostras de teste. É possível perceber aproximação entre os valores obtidos e os esperados. Na Figura 14 é apresentado a capacidade de aproximação da rede PMC, na qual o erro percentual médio das amostras foi $EPM = 5,243 \pm 4,748 \%$.

A Tabela 2 contém o resumo dos resultados obtidos com cada um dos banco de dados (Teste 1), referentes ao primeiro teste. É apresentado o tipo do banco de dados; o número de neurônios na camada oculta da melhor rede; o Erro Percentual Médio (EPM) e o Desvio Padrão referentes à etapa de teste; e as figuras que contêm as informações da escolha da melhor rede e do desempenho desta, respectivamente.

Tabela 2 – Resumo dos resultados obtidos no Teste 1.

Banco de dados	Rede	Camada oculta	EPM	DP	Figuras
MP3	Rede 1	34 neurônios	4,398	3,887	6 e 8
WAV	Rede 2	31 neurônios	5,293	4,215	9 e 11
Misto	Rede 3	38 neurônios	5,243	4,748	12 e 14

Em todos os treinamento realizados (Figuras 6, 9 e 12) é possível perceber

Figura 12 – EPM das redes PMC, para o banco misto, em função do número de neurônios na camada oculta.

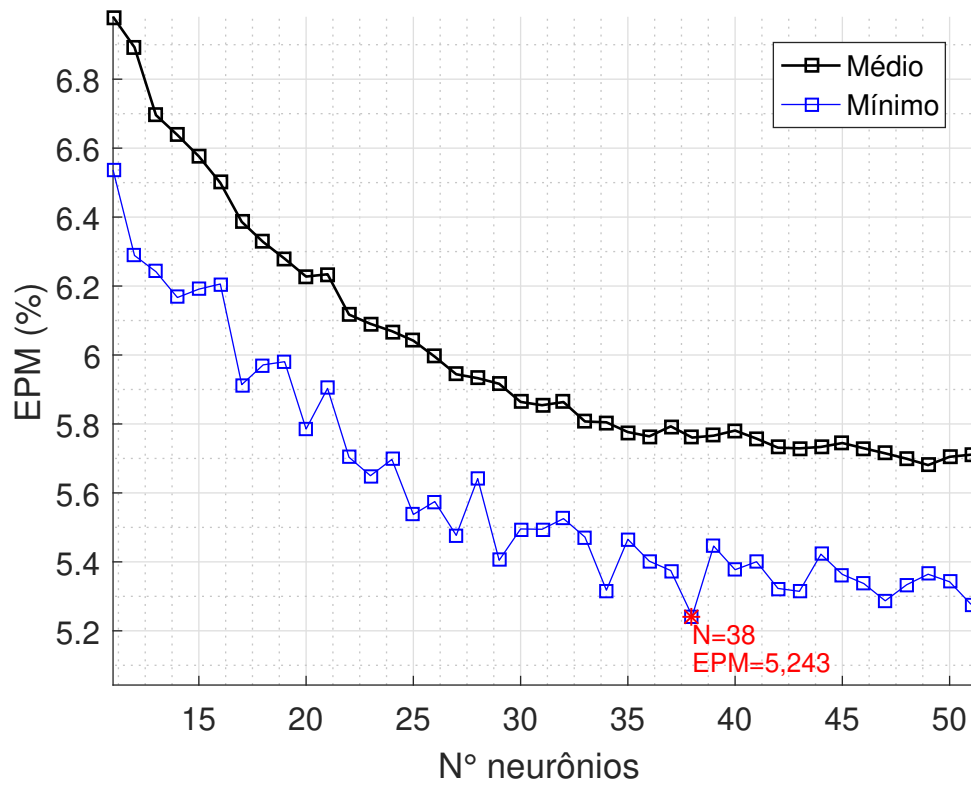


Figura 13 – Saída estimada para os dados de teste pela rede com o menor EPM (Rede 3).

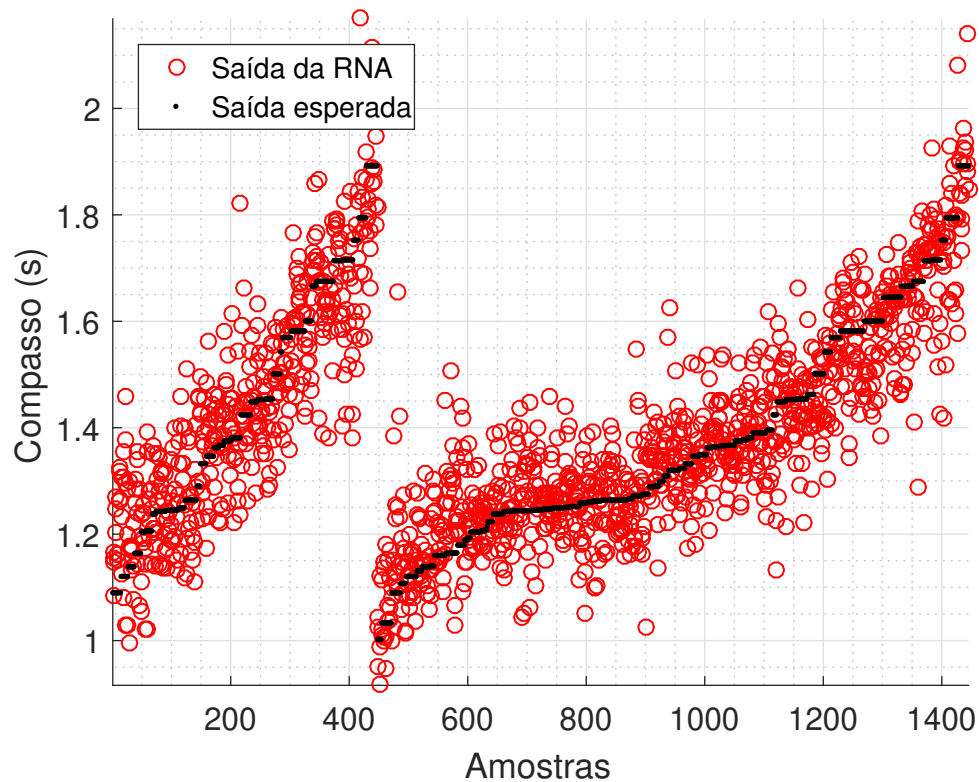
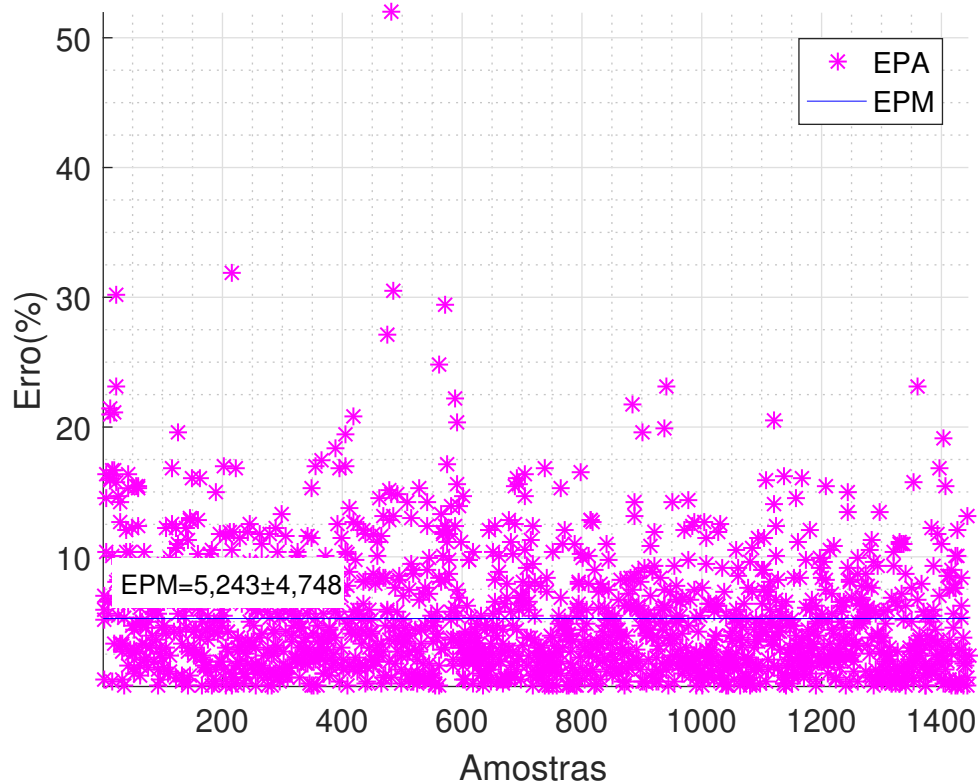


Figura 14 – Erro obtido com os dados de teste pela rede com o menor EPM (Rede 3), referente à Figura 13.



Fonte: Elaborado pelo autor no Matlab.

distanciamento entre a média dos EPM para os 100 treinamentos e mínimo EPM. Esse fenômeno é esperado pois os valores de pesos sinápticos e potenciais de ativação são iniciados com valores aleatórios que podem levar o treinamento cessar em mínimos erros locais, isso justifica muitas repetições de treinamentos para cada configuração de rede para o sorteio de novos pesos. Quanto maior o número de treinamentos com a mesma configuração de rede e novos sorteios dos pesos, maior a chance de serem sorteados valores iniciais que favoreçam o aprendizado da rede.

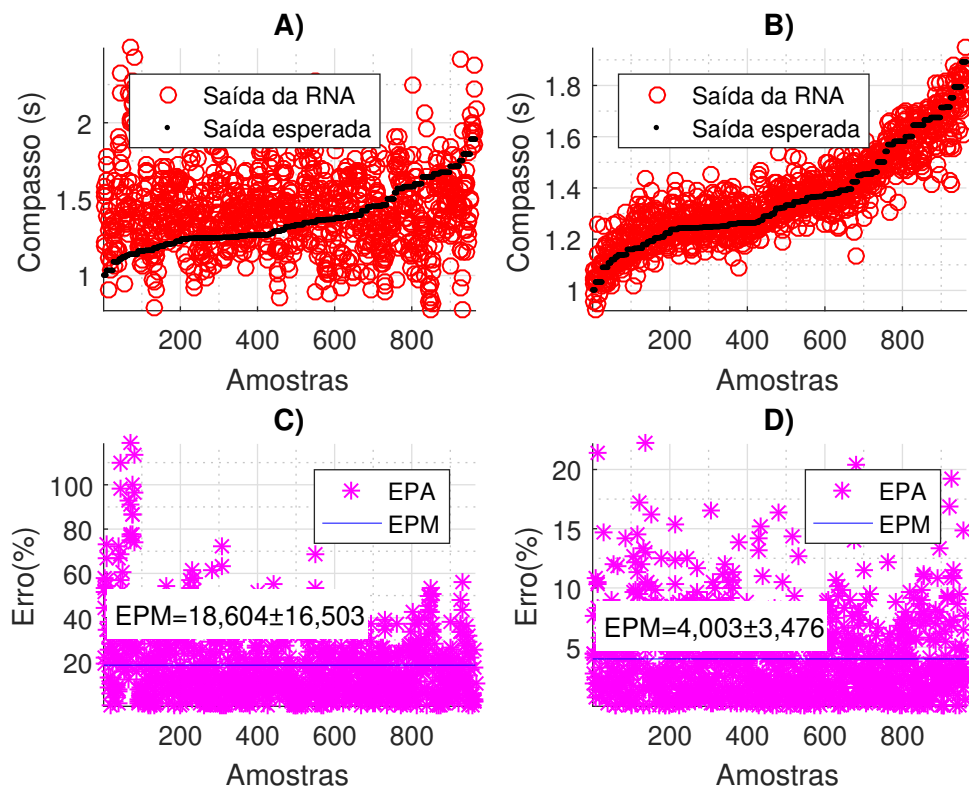
O aspecto de “escada” representado pelos pontos pretos nas Figuras 7, 10 e 13, onde em um mesmo degrau estão as amostras de uma mesma música, se deve a ordenação dos segmentos das músicas dos compassos menores para os maiores (das músicas mais rápidas para as mais lentas). Podemos observar que a distância entre os valores esperados (em pontos pretos) e encontrados (em círculos vermelhos) varia entre as músicas e também entre os segmentos de uma mesma música. Os Desvios Padrões (DP) apresentados nas Figuras 8, 11 e 13 podem ser explicados pela diferença entre as músicas, devido aos diferentes compassos atribuídos a cada uma delas. A diferença entre os segmentos de uma mesma música ocorre devido a diferença natural de componente espectral existente entre um trecho e outro da música e a diferença do ritmo dentro de uma mesma música que pode existir, mas foi desconsiderada para simplificação do trabalho.

Os resultados encontrados mostraram que é possível estimar o tempo de execução do passo base de músicas de forró a partir do espectro de frequências da música utilizando uma rede PMC com uma única camada oculta os três treinamentos encontraram redes com acurácia acima de 94%. Todas as melhores redes obtidas possuem entre 30 e 40 neurônios na camada oculta. A simplicidade dos modelos encontrados facilita a implementação destes em um aplicativo para dispositivos móveis.

4.4 Comparação das três abordagens

Na Figura 15 é apresentado o desempenho das Redes 2 e 3 no Teste 2, ou seja, para um banco de amostras de músicas em MP3. O desempenho da Rede 1 foi apresentado na Figura 8, pois o banco de Teste 2 é o mesmo do banco em MP3 do Teste 1. A Rede 2 teve dificuldade em estimar o valor do compasso com $EPM = 18,604 \pm 16,503 \%$ diferentemente da Rede 3 que obteve $EPM = 4,003 \pm 3,476 \%$.

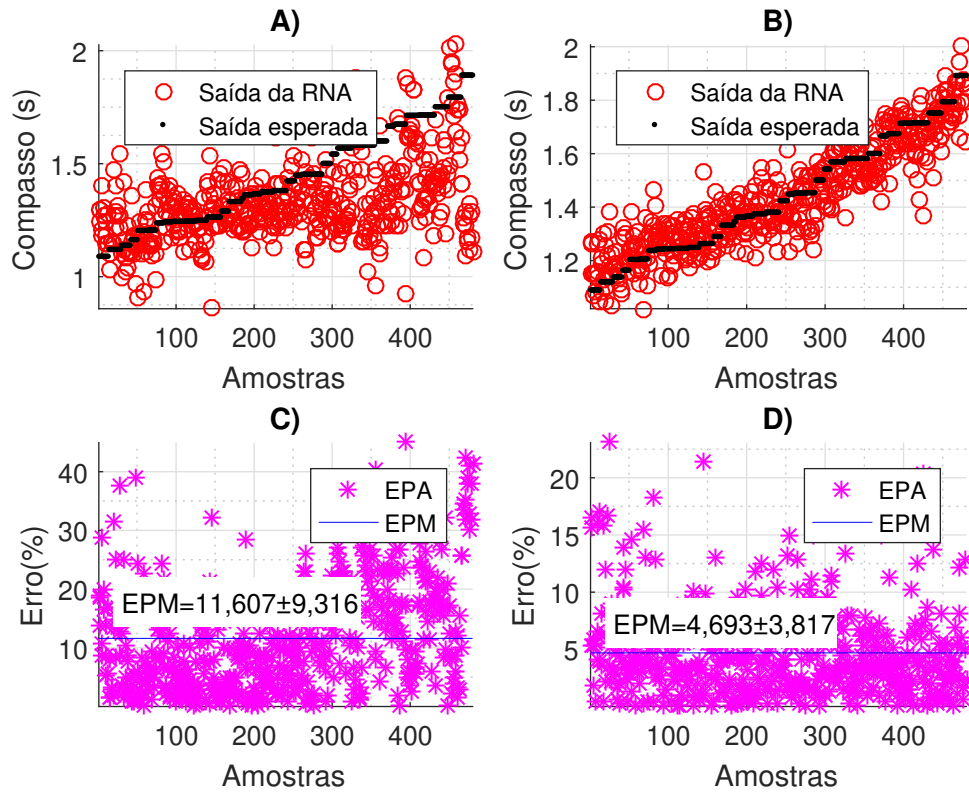
Figura 15 – Desempenho das Redes 2 e 3 no Teste 2 (Banco em MP3): A) Valores estimados pela Rede 2. B) Erro percentual da Rede 2. C) Valores estimados pela Rede 3. D) Erro percentual da Rede 3.



Na Figura 16 é apresentado o desempenho das Redes 1 e 3 no Teste 3, ou seja, para um banco de amostras de músicas em WAV. O desempenho da Rede 2 foi apresentado na Figura 11, pois o banco de Teste 3 é o mesmo do banco em WAV do Teste 1. A Rede

1 teve dificuldade em estimar o valor do compasso com $EPM = 11,607 \pm 19,316 \%$ diferentemente da Rede 3 que obteve $EPM = 4,693 \pm 3,817 \%$.

Figura 16 – Desempenho das Redes 1 e 3 no Teste 3 (Banco em WAV): A) Valores estimados pela Rede 1. B) Erro percentual da Rede 1. C) Valores estimados pela Rede 3. D) Erro percentual da Rede 3.



Fonte: Elaborado pelo autor no Matlab.

A Tabela 3, sintetiza os resultados obtidos pelas três redes selecionadas no Teste 1 quando submetidas aos Testes 2 e 3. Nos resultados encontrados, a Rede 3 com $EPM = 4,003 \pm 3,476 \%$ no Teste 2 e $EPM = 4,693 \pm 3,817 \%$ no Teste 3, obteve o menor EPM e DP que as Redes 1 e 2 nos dois testes. Portanto, a Rede 3 com 38 neurônios na camada oculta e treinada a partir do banco de dado misto é a melhor rede para estimar a duração do compasso das músicas em WAV e MP3. No entanto, ao investigar as causas do melhor desempenho da Rede 3, percebeu-se que cerca de 70% dos dados apresentados nos Testes 2 e 3 fizeram parte do treinamento desta rede. Este fato, coloca em cheque a neutralidade dos Testes 2 e 3 e conseqüentemente, a superioridade da Rede 3 em ambos os testes. A decisão de reutilizar os bancos do Teste 1 para os Testes 2 e 3, se deu para aproveitar o banco de dados. Para separar bancos de teste inéditos para a segunda fase de teste (Testes 1 e 2) será necessário ampliar o banco de dados utilizado.

O tempo médio do compasso do banco de dados é 1,290 s, este valor corresponde ao compasso da música “Baioneiro Gonzagão” (Tabela 1). A Rede 3, obteve $EPM = 4,693\%$ para músicas em WAV, para está música o erro corresponde a aproximadamente 60 ms.

Tabela 3 – Resumo dos resultados obtidos no Testes 2 e 3.

Rede	Banco	Teste 2		Banco	Teste 3	
		Resultado	Figuras		Resultado	Figuras
1	MP3	$4,398 \pm 3,887$	8	WAV	$11,607 \pm 9,316$	16 C)
2	MP3	$18,604 \pm 16,503$	15 C)	WAV	$5,293 \pm 4,215$	11
3	MP3	$4,003 \pm 3,476$	15 D)	WAV	$4,693 \pm 3,817$	16 D)

Desta forma, o modelo proposto, representado pela Rede 3, estima a duração do compasso com erros médios em torno de ± 60 ms.

Os erros obtidos pela Rede 3, sustentam a hipótese de que com o treinamento a partir de banco de dados misto, músicas em MP3 sem ruído e músicas em WAV ruidosas, capacita a rede neural a valorizar as componentes mais significativas do espectro da música. Em contrapartida, o baixo desempenho da Rede 1 quando apresentada ao conjunto de músicas ruidosas (Teste 2) mostra a necessidade de acrescentar entradas ruidosas no conjunto de dados de treinamento, para se obter acurácia na estimação da duração do compasso de músicas sujeitas a ruído.

Neste trabalho o ruído não foi adicionado artificialmente, como as músicas com ruído (gravadas em WAV) foram gravadas em um espaço tradicional de dança, os ruídos acrescentado são da mesma natureza que os ruídos que serão encontrados pela rede quando implementada for em um aplicativo móvel. Trabalhos futuros com análise da natureza desse ruído poderão permitir o tratamento dos ruídos antes da rede neural ou colocá-los artificialmente na música de forma a garantir a generalização da rede neural, sem o exaustivo trabalho de regravar as músicas em um espaço de dança.

5 Considerações Finais

Este trabalho propôs a utilização de uma rede PMC para auxiliar surdos na interpretação de músicas de forró. Os resultados encontrados mostraram que é possível estimar o tempo de execução do passo base de músicas de forró a partir do espectro de frequências da música, utilizando uma rede PMC com uma única camada oculta. A rede neural treinada com um banco de dados de músicas com e sem ruídos foi a melhor para os dois cenários com erro percentual médio entre 5% e 6% permitindo que pessoas surdas consigam sentir a música com erros médios entorno de ± 60 ms. A simplicidade do sistema proposto favorece a sua implementação em um aplicativo para dispositivo móvel. As principais limitações do trabalho se dão pelo processo manual de estimação da duração do compasso esperado que desconsidera a variação de ritmo dentro de uma mesma música. Será necessário aumentar o banco de dados para separar uma parte do banco exclusiva para a escolha da melhor rede. A partir dos resultados encontrados, o próximo passo é desenvolver um aplicativo que transforme sinais sonoros em outros estímulos sensoriais, mostrando o tempo de marcação da música ao usuário. A partir da implementação do aplicativo será possível avaliar se a acurácia obtida pela rede é suficiente para fornecer ao surdo uma experiência significativa.

Referências

- BHALKE, D. G.; RAO, C. B.; BORMANE, D. S. Automatic musical instrument classification using fractional fourier transform based- MFCC features and counter propagation neural network. *Journal of Intelligent Information Systems*, Journal of Intelligent Information Systems, v. 46, n. 3, p. 425–446, 2016. ISSN 15737675. Disponível em: <<http://dx.doi.org/10.1007/s10844-015-0360-9>>. Citado na página 25.
- BOSSEY, A. Accessibility all areas? UK live music industry perceptions of current practice and Information and Communication Technology improvements to accessibility for music festival attendees who are deaf or disabled. *International Journal of Event and Festival Management*, v. 11, n. 1, p. 6–25, feb 2020. ISSN 1758-2954. Disponível em: <<https://www.emerald.com/insight/content/doi/10.1108/IJEFM-03-2019-0022/full/html>>. Citado na página 13.
- BRASIL. Lei nº. 10.260, de 12 de jul. de 2001. *Dispõe sobre o Fundo de Financiamento ao estudante do Ensino Superior e dá outras providências*, Brasília, DF, Jul. 2001. Citado na página 12.
- BRASIL. Lei nº. 11.096, de 13 de jan. de 2005. *Institui o Programa Universidade para Todos - PROUNI*, Brasília, DF, Jan. 2005. Citado na página 12.
- BRASIL. Decreto nº. 6.096, de 24 de abr. de 2007. *Institui o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais - REUNI*, Brasília, DF, Abr. 2007. Citado na página 12.
- BRASIL. Decreto nº. 7.234, de 19 de jul. de 2010. *Dispõe sobre o Programa Nacional de Assistência Estudantil - PNAES*, Brasília, DF, Jul. 2010. Citado na página 12.
- BRASIL. Lei nº. 12.688, de 18 de jul. de 2012. *Institui o Programa de Estímulo à Reestruturação e ao Fortalecimento das Instituições de Ensino Superior (Proies)*, Brasília, DF, Jul. 2012. Citado na página 12.
- BRASIL. Lei nº. 12.711, de 29 de ago. de 2012. *Dispõe sobre o ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências*, Brasília, DF, Ago. 2012. Citado na página 12.
- BRASIL. Lei nº. 12.711, de 29 de ago. de 2012. *Dispõe sobre o ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências*, Brasília, DF, Ago. 2012. Citado na página 12.
- BRASIL. Lei nº. 13.409, de 28 de dez. de 2016. *Reserva de Vagas para Pessoas com Deficiência nos Cursos Técnicos de Nível Médio e Superior das Instituições Federais de Ensino*, Brasília, DF, Dez. 2016. Citado na página 12.
- CAI, L.; CAI, Q. Music creation and emotional recognition using neural network analysis. *Journal of Ambient Intelligence and Humanized Computing*, Springer Berlin Heidelberg, n. 0123456789, 2019. ISSN 18685145. Disponível em: <<https://doi.org/10.1007/s12652-019-01614-6>>. Citado na página 13.

COOLEY, J. W.; TUKEY, J. W. An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, v. 19, n. 90, p. 297, apr 1965. ISSN 00255718. Disponível em: <<https://www.jstor.org/stable/2003354?origin=crossref>>. Citado na página 17.

GIFFONI, A. *Música brasileira para contrabaixo*. [S.l.]: Irmãos Vitale, 1997. Citado na página 15.

HAGUIARA-CERVELLINI, N. *A musicalidade do surdo: representação e estigma*. [S.l.]: Plexus Editora, 2003. Citado na página 12.

HAYKIN, S. *Redes neurais: princípios e prática*. [S.l.]: Bookman Editora, 2007. Citado 2 vezes nas páginas 18 e 19.

JUNIOR, A. C. d. Q.; VOLP, C. M. Forro Universitário: a tradução do forro nordestino no sudeste brasileiro. *Motriz-revista De Educacao Fisica*, v. 11, n. 2, p. 117–120, 2005. ISSN 1980-6574. Citado 2 vezes nas páginas 15 e 17.

JUNIOR, A. C. de Q. et al. Caracterização do Xote e do Baião dançados no interior do Estado de São Paulo. *Movimento (ESEFID/UFRGS)*, v. 15, n. 3, p. 233–247, mar 2009. ISSN 1982-8918. Disponível em: <<https://seer.ufrgs.br/Movimento/article/view/2347>>. Citado 2 vezes nas páginas 15 e 16.

LACERDA, C. B. F. de. A inclusão escolar de alunos surdos: o que dizem alunos, professores e intérpretes sobre esta experiência. *Cadernos CEDES*, v. 26, n. 69, p. 163–184, 2006. Citado na página 12.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity*. *Bulletin of Mathematical Biology*, v. 52, n. 1, p. 99–115, 1990. Citado na página 18.

MIRZAEI, M.; KÁN, P.; KAUFMANN, H. EarVR: Using Ear Haptics in Virtual Reality for Deaf and Hard-of-Hearing People. *IEEE Transactions on Visualization and Computer Graphics*, v. 26, n. 5, p. 2084–2093, 2020. ISSN 19410506. Citado na página 13.

MORÉ, J. J. The levenberg-marquardt algorithm: implementation and theory. In: *Numerical analysis*. [S.l.]: Springer, 1978. p. 105–116. Citado na página 19.

NUNES, E.-b. V. M.; SOUZA, S. F. D. O Forró E Suas Configurações: a Aliteração, a Mulher, O Homem, O Movimento Corporal E O Ambiente Nas Canções Forrozísticas. *Revista Entrelinhas*, v. 8, n. 1, p. 29, 2014. ISSN 1806-9509. Citado 2 vezes nas páginas 15 e 17.

OPPENHEIM, A. V.; WILLSKY, A. S. *Sinais e Sistemas, 2. ed.* [S.l.: s.n.], 2010. 592 p. ISBN 9788543013800. Citado na página 17.

PAULA, T. R. M. de; PEDERIVA, P. L. M. Musical Experience in Deaf Culture. *International Journal of Technology and Inclusive Education*, 2017. Disponível em: <<https://www.semanticscholar.org/paper/Musical-Experience-in-Deaf-Culture-Paula-Pederiva/1e450c197d715f94ee8b062a5eb85f26c33adf5e>>. Citado na página 12.

RISTOFF, D. O Novo Perfil do Campus Brasileiro. *Avaliação (Campinas) [online]*, v. 19, n. 3, p. 723–747, 2014. Citado na página 12.

- SALAMON, J.; BELLO, J. P. Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *IEEE Signal Processing Letters*, v. 24, n. 3, p. 279–283, 2017. ISSN 10709908. Citado na página 13.
- SCHOENBERG, A. *Fundamentos da composição musical*. [S.l.]: Edusp, 1990. v. 53. 1–136 p. ISSN 1098-6596. Citado na página 15.
- SHARP, A.; BACON, B. A.; CHAMPOUX, F. Enhanced tactile identification of musical emotion in the deaf. *Experimental Brain Research*, Springer Berlin Heidelberg, v. 238, n. 5, p. 1229–1236, 2020. ISSN 14321106. Disponível em: <<https://doi.org/10.1007/s00221-020-05789-9>>. Citado na página 13.
- SHARP, A. et al. Musicians show better auditory and tactile identification of emotions in music. *Frontiers in Psychology*, v. 10, n. AUG, p. 1–7, 2019. ISSN 16641078. Citado na página 13.
- SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. *Redes Neurais Artificiais para Engenharia e Ciências Aplicadas: Fundamentos Teóricos e Aspectos Práticos*. 2ª edição. ed. São Paulo: Artliber, 2016. Citado 4 vezes nas páginas 13, 19, 25 e 26.
- SOLANKI, A.; PANDEY, S. Music instrument recognition using deep convolutional neural networks. *International Journal of Information Technology*, Springer Singapore, 2019. ISSN 2511-2104. Disponível em: <<https://doi.org/10.1007/s41870-019-00285-y>>. Citado na página 13.
- TITZE, I. R.; MARTIN, D. W. Principles of voice production. *The Journal of the Acoustical Society of America*, v. 104, n. 3, p. 1148–1148, 1998. ISSN 0001-4966. Citado na página 16.
- WHITE, R. D.; GROSH, K. Microengineered hydromechanical cochlear model. *Proceedings of the National Academy of Sciences of the United States of America*, v. 102, n. 5, p. 1296–1301, 2005. ISSN 00278424. Citado na página 17.
- YU, D. et al. Predominant Instrument Recognition Based on Deep Neural Network with Auxiliary Classification. *IEEE/ACM Transactions on Audio Speech and Language Processing*, IEEE, v. 28, p. 852–861, 2020. ISSN 23299304. Disponível em: <<https://ieeexplore.ieee.org/document/8979336/>>. Citado 2 vezes nas páginas 13 e 25.
- ZITOMER, M. R. Always being on your toes: elementary school dance teachers' perceptions of inclusion and their roles in creating inclusive dance education environments. *International Journal of Inclusive Education*, v. 21, n. 4, p. 428–440, 2017. ISSN 14645173. Citado na página 14.