

Estimação do compasso musical do forró utilizando rede perceptron multicamadas

Lucas F. Paiva* Hugo G. Lopes* Leonardo B. Felix*
Rodolpho V. A. Neves*

* *Departamento de Engenharia Elétrica, Universidade Federal de Viçosa, MG. (e-mail: lucas.f.paiva@ufv.br, hugolopg@gmail.com, leobonato@ufv.br e rodolpho.neves@ufv.br).*

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 “forró”, an activity to gather students with their peers. The “forró” is a musical genre marked by percussion instruments, such as the triangle and the zabumba, and by musical compass that dictates the movement speed of the dance. Aiming the inclusion of deaf people in cultural activities involving “forró”, this paper proposes to estimate the musical compass time for “forró” musics using a multilayer perceptron neural network (MLP). A “forró” music database with “Forró Pé-de-serra” and “Forró 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 “forró” instructor, that danced the musics and measured the time to perform the fundamental steps. The musical compass times estimated by the MLP reached 96% of accuracy (percent mean error of 4%), allowing deaf people to feel the music with a mean error of ± 65 ms. The applicability of the approach will be addressed in future works.

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 precisão de 96% (erro médio percentual aproximado de 4%), permitindo que pessoas surdas consigam sentir a música com erros médios de ± 65 ms. A aplicabilidade da abordagem será tratada em trabalhos futuros.

Keywords: Multilayer perceptron; Parameter estimation; Deaf inclusion; Forró; Dance.

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

1. INTRODUÇÃO

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 (Haguiara-Cervellini, 2003), alunos surdos participam de projetos de dança. Entretanto, a partir da vigência da Lei N° 13.409 (2016) (Brasil, 2016),

que altera a “Lei das Cotas” (Lei N° 12.711, 2012) (Brasil, 2012), a tendência é que alunos surdos estejam mais presentes nestes ambientes. 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. Alguns trabalhos mostram que é possível potencializar o contato do surdo com a música através de estímulos visuais e táteis (Lopes e Araújo,

2009), (Luiz e Araújo, 2008) e (Luiz, 2008). Entretanto, para que a música possa ser convertida de uma onda sonora para estímulos sensoriais, não próximas aos alto-falantes, é necessário alguma ferramenta de interpretação, como uma rede neural artificial (RNA).

A rede neural artificial é um algoritmo computacional treinado para classificar dados, aproximar funções e fazer previsões (Silva et al., 2016). A literatura sobre processamento de áudio e reconhecimento de padrões utilizando aprendizado de máquinas é extensa, podendo ser citados trabalhos como Adams et al. (2006), Patil e Nemade (2016), Murthy et al. (2017) e Finley e Razi (2019). 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.

Sendo assim, o objetivo deste trabalho é estimar a duração de um compasso musical e definir a duração do passo base da dança fornecendo a velocidade que a pessoa deve dançar a música. Uma rede perceptron multicamadas (PMC) deve estimar o tempo correto de execução de um movimento do passo base da dança. O pré-processamento das músicas, o treinamento da rede PMC e a avaliação do resultado foram realizados no Matlab, utilizando um banco de dados de músicas variadas de forró.

2. FORRÓ E SUAS CARACTERÍSTICAS

A acepção do termo “forró” é múltipla, segundo de Quadros Junior e Volp (2005), não existe um consenso sobre significado do termo, destaca ainda que talvez este consenso nunca ocorra. Em diálogos referindo-se ao forró, a palavra pode significar uma festa ou local (Rocha, 2004) e um gênero musical (Giffoni, 1997). Neste trabalho são consideradas três componentes do fenômeno forró: as músicas que são tocadas, a forma de se dançar e a festa que reúne as pessoas para ouvir e dançar xote, baião, xaxado e rasta pé (Nunes e de Souza, 2014).

O forró, enquanto gênero musical, pode ser classificado em três categorias: Forró Pé-de-serra, Forró Universitário e Forró Eletrônico (de Quadros Junior e 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. Entretanto, o Forró Eletrônico não é muito tocado no contexto universitário.

2.1 Sinal sonoro

Uma onda é caracterizada por sua amplitude (ou energia) e frequência em que oscila. A música percebida pelos ouvidos humanos podem ser caracterizadas pela composição de várias ondas sonoras, com 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 e Grosh, 2005). A partir dessa decomposição que ocorre a diferenciação entre os sons graves dos agudos, mesmo quando enviados ao mesmo tempo.

Computacionalmente, a Transformada Rápida de Fourier (FFT – Fast Fourier Transform) é um algoritmo matemático eficiente utilizado para a obtenção das amplitudes e fases dos diversos componentes do sinal estudado (Cooley e Tukey, 1965; Anders, 1995).

2.2 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 e 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 (Nunes e de Souza, 2014; de Quadros Junior e Volp, 2005). 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.

2.3 Passo Base Frente e Trás

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 (de Quadros Junior et al., 2009).

A Figura 1 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 de Quadros 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 1, 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, ...).

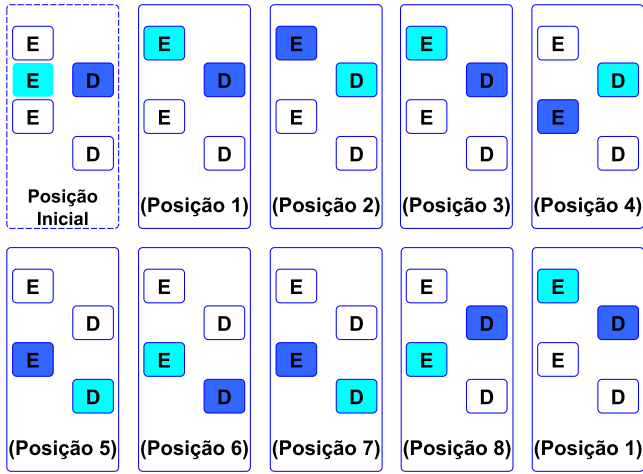


Figura 1. Diagrama do Passo Base Frente e Trás. O pé esquerdo (E) e direito (D) são representados pelos quadrados com as letras e as cores azul escuro e azul claro representam o pé de sustentação e o pé que toca o chão, respectivamente.

2.4 Rede Perceptron Multicamadas

Uma rede neural é um algoritmo que interpreta informações de sinais de entrada e entrega uma saída significativa ao usuário. A Figura 2 apresenta a estrutura de uma rede neural com uma camada oculta.

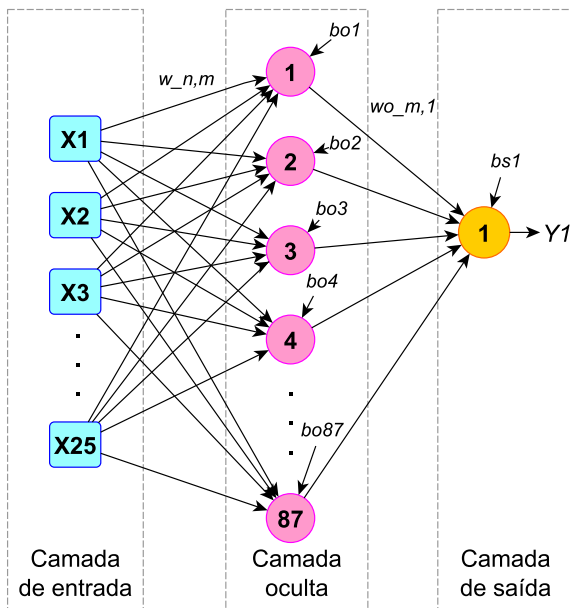


Figura 2. Estrutura de uma rede PCM com uma camada oculta.

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 $b_{o1}, b_{o2}, b_{o3}, \dots, b_{om}$ e b_{s1} 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.

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 et al., 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 et al., 2016). Para estimar o tempo de duração de um compasso, as entradas foram as amplitudes das frequências do som mais grave da zabumba e a saída a duração do compasso aferido pela duração da execução de um passo base.

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 et al., 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. MATERIAL E MÉTODOS

As músicas selecionadas para a realização deste trabalho são do estilo Forró Pé-de-serra e Forró Universitário. Foram selecionadas 40 músicas para a compor o banco de dados. 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.

A seleção das músicas é apresentada na Tabela 1, com seus respectivos nomes e artistas ou bandas responsáveis pelas gravações das versões utilizadas. A extração das características das músicas, o treinamento da PMC e a avaliação do resultado obtido foram realizados no Matlab.

3.1 Banco de dados

A preparação das músicas para o banco de dados consistiu em retirar o início e o final das músicas e normalizar

Tabela 1. Músicas selecionadas para o banco de dados.

Música	Banda/Artista
Avisa	Falamansa
100 anos	Falamansa
Acalanto	Black e Bárbara Greco
Alambique de Barro	Os 3 do Nordeste
Amanhecer	Os 3 Cangaceiros
Anjo da Guarda Noturno	Bicho de Pé
Asas	Bicho de Pé
Baião Caçula	Dominguinhos
Baião de Respeito	Dona Zaira
Baião Jovem	Trio Cangaço
Baioneiro Gonzagão	Trio Dona Zefa
Luiz Gonzaga do Nascimento	Balança Eu
Bata a Porta	Black
Briga de Artista	Trio Dona Zefa
Canto do Sabiá	Trio Dona Zefa
Casinha Pronta	Black
Cigano Terrorista	Trio Cangaço
Como Todo Amor	Forrueiros
De Tamanco Mulher	Trio Forrozão
Desilusão	Dominguinhos
É Proibido Cochilar	Circuladô de Fulô
Encosta n'Eu	Estakazero
Esconderijo do Amor	Genaro
Eu e Você	Bicho de Pé
Feitiço da Lua	Forrueiros
Fino Trato	Forró na Contramão
Flor de Lis	Circuladô de Fulô
Lilith	Alcalyno
Medo de Escuro	Falamansa
Nosso Xote	Bicho de Pé
Nossos Olhos	Forró Massapê
Produto Nacional	Trio Dona Zefa
Seu Siriri	Trio Dona Zefa
Ta Esquentando	Trio Dona Zefa
Tem Problema Não	Trio Bastião
Vem me Ver	Circuladô de Fulô
Vida Boa Danada	Trio Dona Zefa
Você vai se Machucar	Trio Dona Zefa
Xote dos Milagres	Falamansa
Xote Universitário	Falamansa

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 (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 0,5 segundos entre os segmentos.

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

À 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 Estimação das saídas desejadas

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 como referência, ou seja, toda vez que a instrutora pisava com o pé direito a frente, o cronômetro era reiniciado. 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.

3.3 Treinamento da 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.

A topologia de rede escolhida foi o perceptron multicamadas (PMC), devido sua característica de aproximação universal de funções (Silva et al., 2016). Outra característica desta rede é necessitar de apenas uma camada oculta para mapear qualquer função contínua no espaço das funções reais, desde que, seja utilizada uma função de ativação contínua e limitada em sua imagem (Silva et al., 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; e,
- 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 100 redes, variando o número de neurônios na camada oculta de 1 a 100. Cada rede foi treinada 100 vezes com novos sorteios dos pesos e potenciais de ativação. 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

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

na qual N é o número de amostras durante o treinamento, Yd_i é a saída desejada e Y_i é a saída estimada da rede PMC. Foi considerada a melhor rede aquela que obteve menor média dos Erro Percentual Médio (EPM) dos 100 treinamentos realizados. O EPM pode ser definido por

$$\text{EPM}(\%) = \frac{1}{N} \sum_{i=1}^N \frac{|Yd_i - Y_i|}{Yd_i} \times 100\%, \quad (3)$$

com Y_d o valor esperado, Y o valor encontrado e N o número de amostras.

4. RESULTADOS

As 40 músicas segmentadas totalizaram 2830 amostras. As durações dos compassos variaram de 1,09 s até 1,89 s, sendo os menores tempos predominantemente relacionados às músicas mais rápidas e os maiores, às músicas mais lentas.

A melhor rede PMC foi obtida utilizando 87 neurônios na camada oculta, obtendo $\text{EPM} = 3,408\%$, para a fase de teste. No entanto, pode-se observar na Figura 3 que para implementação, redes menores podem ser utilizadas com aumento de 0,1% do erro, como o caso da rede de 22 neurônios com $\text{EPM} = 3,503\%$, e a rede com 46 neurônios, que alcançou $\text{EPM} = 3,459\%$. Para implementação da rede PMC, o número de neurônios é um fator decisivo pois aumenta a complexidade da rede e define o esforço computacional necessário para fornecer a saída apropriada. Neste estudo, como o foco é encontrar a rede como menor erro, portanto, a melhor rede foi a constituída por 87 neurônios na camada oculta.

Na Figura 4, é apresentado o comportamento da rede escolhida (87 neurônios na camada oculta) para as amostras de teste. É possível perceber aproximação entre os valores obtidos e os esperados. O aspecto de “escada” representado pelos pontos pretos na Figura 4, 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 preto) e encontrados (em vermelho) varia entre as músicas e também entre os segmentos de uma mesma música. A diferença entre as músicas ocorre pelos diferentes compassos atribuídos a cada uma delas. A diferença entre os segmentos de uma mesma música pode ser devido a dois fatores: a diferença natural existente entre um trecho e outro da música e a diferença do ritmo dentro de uma mesma música, sendo que a todos os segmentos de uma música foram atribuídos o mesmo tempo de compasso.

Na Figura 5 é possível perceber a capacidade de aproximação da rede PMC, na qual o erro percentual médio das amostras foi $\text{EPM} = 3,408 \pm 3,42\%$. Na prática, para a música mais lenta, este erro corresponde a 64,411 ms para baixo ou para cima do valor esperado. Em músicas mais rápidas, este erro corresponde a 37,147 ms.

5. CONCLUSÃO

Este trabalho propôs a utilização de uma rede PMC para auxiliar surdos na interpretação de músicas de forró. Os

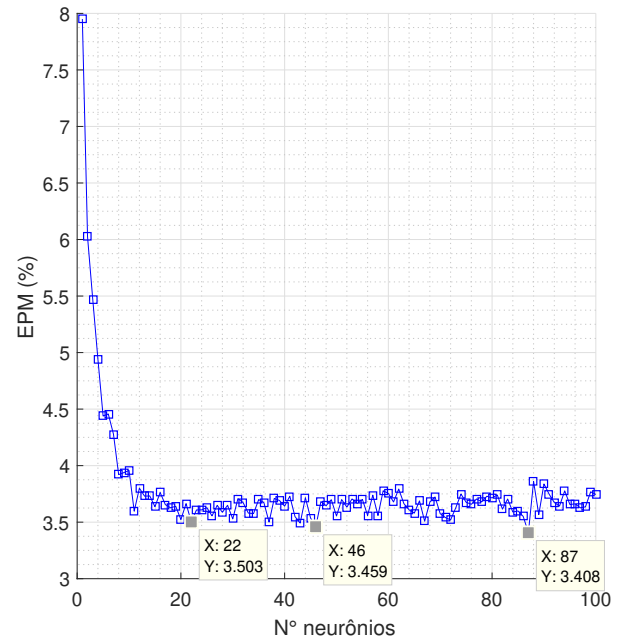


Figura 3. EPM de redes PMC em função do número de neurônios na camada oculta.

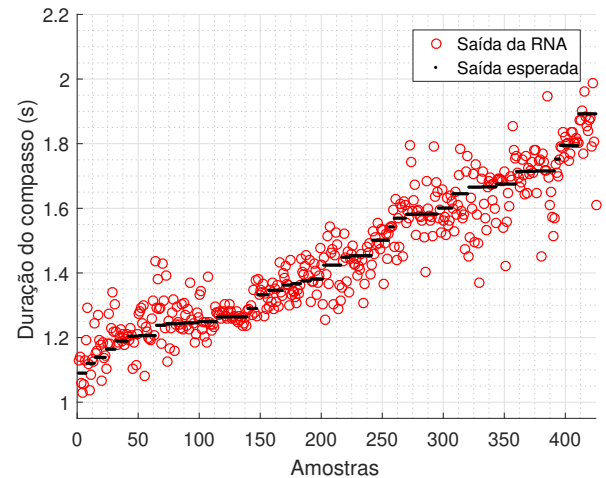


Figura 4. Saída estimada da rede PMC de 87 neurônios na camada oculta para a fase de teste.

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 proposta favorece a implementação de uma ferramenta que demanda baixa capacidade de processamento. Os erros máximos alcançados foram na faixa de 65 ms durante cada compasso estimado. A partir dos resultados encontrados, o próximo passo é desenvolver uma ferramenta que transforme sinais sonoros em outros estímulos sensoriais, mostrando o tempo de marcação da música ao usuário. Como não foi objetivo do trabalho observar se este erro pode comprometer a dança, estudos futuros poderão ser realizados para avaliar o desempenho alcançado.

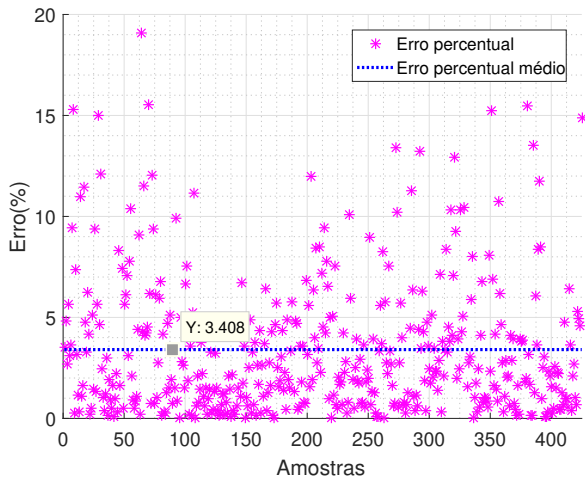


Figura 5. Erro obtido pela rede PMC de 87 neurônios na camada escondida na fase de teste, referente à Figura 4.

REFERÊNCIAS

- Adams, N.H., Bartsch, M.A., e Wakefield, G.H. (2006). Note segmentation and quantization for music information retrieval. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1), 131–141.
- Anders, E.Z. (1995). Understanding the fft: a tutorial on the algorithm and software for laymen, students, technicians and working engineers.
- Brasil (2012). 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.*
- Brasil (2016). 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.*
- Cooley, J.W. e Tukey, J.W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90), 297–301.
- de Quadros Junior, A.C., Fontes, E.C., Dias, R., e Volp, C.M. (2009). Caracterização do xote e do baião dançados no interior do estado de são paulo. *Movimento*, 15(3), 233–247.
- de Quadros Junior, A.C. e Volp, C.M. (2005). Forró universitário: a tradução do forró nordestino no sudeste brasileiro. *Motriz. Journal of Physical Education. UNESP*, 117–120.
- Finley, M. e Razi, A. (2019). Musical key estimation with unsupervised pattern recognition. In *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, 0401–0408. Las Vegas.
- Giffoni, A. (1997). *Música brasileira para contrabaixo*. Irmãos Vitale.
- Haguiara-Cervellini, N. (2003). *A musicalidade do surdo: representação e estigma*. Plexus Editora.
- Lopes, K.F. e Araújo, P.F. (2009). Proposta de ensino de sapateado para crianças surdas. *Revista Brasileira de Ciência e Movimento*, 17(1).
- Luiz, T.R.B. (2008). *O uso de softwares para estimulação da percepção do surdo frente aos parâmetros de velocidade do ritmo: proposta de utilização do Bpm Counter e do Vpm Counter no programa de atividade rítmicas adaptado as pessoas surdas*. Ph.D. thesis, Universidade de Campinas, Campinas.
- Luiz, T.R.B. e Araújo, P.F. (2008). Avaliação de um programa de atividade rítmica adaptada para variação dos parâmetros de velocidade do ritmo para pessoas surdas. *Revista brasileira de Ciência e Movimento*, 11(3), 27–32.
- Murthy, Y.V.S., Koolagudi, S.G., e Swaroop, V.G. (2017). Vocal and non-vocal segmentation based on the analysis of formant structure. In *2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR)*, 1–6. Bangalore.
- Nunes, E.B.V.M. e de Souza, S.F. (2014). O forró e suas configurações: a aliteração, a mulher, o homem, o movimento corporal e o ambiente nas canções forrozísticas. *Entrelinhas*, 8(1), 29.
- Patil, N.M. e Nemade, M.U. (2016). Content-based audio classification and retrieval: A novel approach. In *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICCC)*, 599–606. Jalgaon.
- Rocha, J.M.T. (2004). Forró eletrônico, forró universitário. *FESTIVAL DO FOLCLORE*, 40, 62–71.
- Schoenberg, A. (1990). *Fundamentos da composição musical*. Edusp.
- Silva, I.N., Spatti, D.H., e Flauzino, R.A. (2016). *Redes Neurais Artificiais para Engenharia e Ciências Aplicadas: Fundamentos Teóricos e Aspectos Práticos*. Artliber, São Paulo, 2ª edição edition.
- Titze, I.R. e Martin, D.W. (1998). Principles of voice production.
- White, R.D. e Grosh, K. (2005). Microengineered hydro-mechanical cochlear model. *Proceedings of the National Academy of Sciences*, 102(5), 1296–1301.