Uso de engenharia de características para predição da duração do compasso musical do forró

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

* *Núcleo Interdisciplinar de Análise de Sinais (NIAS)*
Departamento de Engenharia Elétrica
Universidade Federal de Viçosa, MG
e-mail: {hugo.lopes, lucas.f.paiva, rodolpho.neves, leobonato}@ufv.br

Abstract: For humans, dancing is highly associated with musicality and the rhythm of songs, which implies that dance participants need to hear the music in order to interact with each other. However, deaf or hearing impaired people (D/HH) are often excluded from this type of social interaction environment, making it even more difficult for them to live in society. Thus, it is necessary to use techniques such as an ANN to convert sound signals into others that D/HH can perceive. This technique already exists in the literature, using the Fourier transform in part of the frequency spectrum of the music as input of an ANN to determine the base step tempo of forró music, however, it was not analyzed how the variations in the network input affect its performance. Therefore, this work proposes the variation of the number of inputs and the frequency spectrum to determine the best extraction of music characteristics. With this, an extraction of audio features was defined that reduced the error by more than 20% compared to the error of the technique existing in the literature that can be used for application in a tool capable of inserting the deaf and hearing impaired in the forró dance environment, providing a better coexistence of these individuals in the local society.

Resumo: A dança, para os seres humanos, está altamente associada a musicalidade e ao ritmo das canções, o que implica que os participantes da dança precisam ouvir a música para interagir entre si. Entretanto, pessoas surdas ou com alguma deficiência auditiva (S/DA) são muitas vezes excluídas deste tipo de ambiente de interação social, dificultando ainda mais a convivência delas na sociedade. Logo, se faz necessário o uso de técnicas como uma Rede Neural Artificial (RNA) para converter sinais sonoros em outros que S/DA possam perceber. Esta técnica já existe na literatura, utilizando a transformada de Fourier em parte do espectro de frequência da música como entrada de uma RNA para determinar o tempo do passo base de músicas de forró, porém, não foi analisado como a seleção de características da rede afetam o erro da mesma. Portanto, este trabalho propõe a variação do número de entradas e do espectro de frequência para determinar a melhor extração de características da música. Com isso, definiu-se uma extração de características de áudio que reduziu o erro em mais de 20% em comparação com o erro da técnica existente na literatura que pode ser utilizada para a aplicação em uma ferramenta capaz de inserir surdos e deficientes auditivos no ambiente de dança de forró, proporcionando uma melhor convivência destes indivíduos na sociedade local.

Keywords: Forró; Artificial neural network; Deaf inclusion; Dance; Multilayer perceptron.

Palavras-chaves: Forró; Rede neural artificial; Inclusão de surdos; Dança; Perceptron Multicamadas.

1. INTRODUÇÃO

As pessoas com deficiência auditiva no Brasil somaram 9,7 milhões, segundo dados do Censo de 2010 do Instituto Brasileiro de Geografia e Estatística (IBGE, 2012). No entanto, é comum que os surdos encontrem dificuldade para se comunicarem, seja com pessoas ouvintes ou surdas, gerando obstáculos que podem ter implicações no avanço social, emocional e cognitivo. À vista disso, a população surda pode ter sua saúde mental afetada (Chaveiro et al., 2014). Vários fatores podem dificultar a inclusão social e limitar a participação desses indivíduos em grupos sociais (Santos et al., 2013). De acordo com Fox et al. (2020), universitários surdos e deficientes auditivos têm uma taxa maior de pensamentos e tentativas de suicídio quando comparado com ouvintes.

No sentido de tornar realidade a inclusão de surdos, foi outorgada a Lei Nº 10.436/2002, que reconhece a Língua Brasileira de Sinais (Libras) como um meio legítimo de comunicação e expressão para os surdos, além do Decreto Nº 5.626/2005 que regulamenta a Lei citada (Brasil,

⋆ O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

2002, 2005). Neste contexto, este trabalho busca somar à inclusão do surdo na sociedade, baseado na comunidade universitária, dialogando com de Lacerda (2006), que diz "além de práticas que se preocupam com a aprendizagem dos conteúdos curriculares, a inclusão do surdo perpassa pela sua inserção na cultura local". Desta forma, entende-se que a dança tem forte presença em muitas culturas pelo mundo e é através dela que se busca uma facilidade para lidar com problemas do dia a dia, mudando a autoestima desde crianças até a terceira idade (FREITAS, 2019).

No Brasil, o forró, uma manifestação artístico-cultural, que vai além da dança e da música, acabou por conquistar todos os segmentos da sociedade com o seu ritmo contagiante e animado, principalmente a comunidade universitária. Existem vários programas que promovem aulas de dança e atividades culturais onde se dá a prática deste estilo de dança, a integração e a troca de conhecimentos entre os alunos. E mesmo sendo reduzido o número de surdos no ambiente universitário, e destes serem estigmatizados como seres não musicais (de Paula and Pederiva, 2017; Haguiara-Cervellini, 2003), alguns alunos surdos participam destes espaços. Na Universidade Federal de Viçosa (UFV), o projeto "Dança nas Moradias" é um exemplo onde, alunos surdos participam da aula de dança de forró e samba de gafieira, oferecidas por alunos do Curso de Dança da UFV para os estudantes residentes em moradias estudantis.

Na literatura atual, trabalhos mostram que é possível potencializar o contato do surdo com a música através de estímulos visuais e táteis (Sharp et al., 2019, 2020; Bossey, 2020; Mirzaei et al., 2020). Assim, foi escolhido uma RNA para que o tempo de duração do passo base fosse estimado a partir da música em tempo real, para que o mesmo possa ser transmitido por um meio que os S/DA possam perceber durante a reprodução da música.

A RNA é um algoritmo computacional que pode ser treinado para classificar dados, aproximar funções e fazer predições (Silva et al., 2016). A literatura sobre processamento de áudio e reconhecimento de padrões utilizando aprendizado de máquinas é extensa, podendo ser citado trabalhos como Salamon and Bello (2017); Solanki and Pandey (2019); Yu et al. (2020); Cai and Cai (2019). Estes trabalhos citados utilizam técnicas de redes neurais artificiais para a classificação de áudio. Porém, em Ferreira-Paiva et al. (2022) é apresentado um modelo de rede que é capaz de estimar o tempo do compasso de uma música de forró, possibilitando a criação de um aplicativo de sinalização do ritmo da música para a inclusão de surdos e deficientes auditivos no ambiente de forró, onde a música pode ser captada em tempo real por um microfone de um dispositivo móvel e, então, o compasso pode ser estimado utilizando uma RNA.

Entretanto, em Ferreira-Paiva et al. (2022) somente a faixa entre (50 à 300) Hz é utilizada e a resolução da transformada rápida de Fourier (FFT) é reduzida para obter 25 entradas. A redução de dados feita da saída da FFT para as entradas da rede pode ter impacto no desempenho da estimação. Embora a RNA de Ferreira-Paiva et al. (2022) estime o compasso com erros próximos a 5%, quando utilizada em músicas sem ruído, o trabalho também mostrou que o modelo possui queda significativa de desempenho quando treinado com músicas sem ruído e testado com músicas com ruído de um espaço de dança (Ferreira-Paiva et al., 2022).

Sendo assim, este trabalho tem como objetivo analisar a geração das entradas de um modelo baseado em RNA, a partir da transformada de Fourier, para estimação do compasso de músicas de forró, com o intuito de reduzir o erro e mitigar a limitação do modelo ao estimar o compasso da música a partir de amostras com ruído de diferentes naturezas. Para isso, foram analisadas maiores números de entradas para a faixa de frequência de 50 Hz à 300 Hz e faixas de frequências mais extensas para o melhor número de entradas encontrado.

2. ESTIMAÇÃO DO COMPASSO DE UMA MÚSICA DE FORRÓ

Para este trabalho, utilizou-se elementos de análise de sinais e redes neurais artificiais para a estimação do ritmo do passo base de forró. Nesta seção, consta a composição do passo base de forró assim como a descrição do tempo necessário para o mesmo ser executado. Em seguida, está apresentado como a literatura lidou com o processamento do sinal da música para a estimação da duração do passo base.

O gênero musical "forró" pode ser classificado em 3 categorias, conforme de Quadros Junior and Volp (2005):

(1) Forró Pé-de-serra: originou-se no Nordeste brasileiro, por volta de 1940, com influência do ambiente rural do sertanejo.
(2) Forró Universitário: tem como origem o Forró Pé-de-serra, porém com influência de outros estilos musicais por jovens sulistas na década de 1990.
(3) Forró Eletrônico: também originado na década de 1990, surge a partir da utilização de instrumentos eletrônicos e com visual chamativo e linguagem estilizada.

O ritmo do forró é passado para o dançarino, principalmente, pelo ritmo da zabumba, onde quatro batidas definem o período de um compasso e dois compassos completam o período de um passo base. Os passos devem ser sincronizados de acordo com as batidas da zabumba (Santos et al., 2018; Schoenberg, 1990).

Neste trabalho, a rede foi treinada para estimar o tempo de duração de um compasso, conforme (Paiva et al., 2020). Entretanto, segundo Schoenberg (1990), a duração de um passo completo consiste no tempo de dois compassos, logo a saída da rede deve ser multiplicada por 2 para sinalizar o passo completo.

2.1 Trabalhos existentes na literatura

Em Paiva et al. (2020) é apresentado uma técnica para estimação de duração do compasso de músicas de forró, onde essa técnica é expandida para músicas com ruído real e ruído branco em Ferreira-Paiva et al. (2022). Para garantir a generalização dos resultados, foram selecionadas músicas com tempos de compasso menores (mais rápidas) e maiores (mais lentas). As músicas foram escolhidas por variedade rítmica e por popularidade no contexto

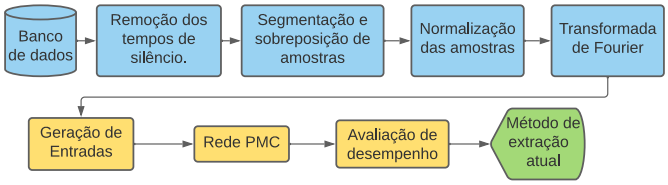


Figura 1. Fluxograma do trabalho realizado em Paiva et al. (2020), onde o banco de dados consiste em 40 músicas sem ruído. O mesmo procedimento foi utilizado em Ferreira-Paiva et al. (2022) para o mesmo banco de dados utilizado neste trabalho.

universitário através da orientação de uma instrutora de dança.

A preparação das músicas para o banco de dados consiste em retirar o início e o final das faixas, segmentar e normalizar a amplitude do sinal em todo domínio do tempo. Cada música foi segmentada em trechos de três segundos com sobreposição de dois segundos entre os segmentos. Em seguida, a normalização foi feita dividindo-se cada valor pelo valor eficaz da amostra V_{rms}, definido por (1), para que eventuais disparidades entre o volume das músicas não interfiram nos resultados. A normalização deve ser feita por amostra em vista que no caso de previsão em tempo real, não haverá dados do futuro da música para a normalização, logo deve-se considerar essa informação inexistente na geração de amostras.

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

À cada segmento foi associado à duração do compasso da música a qual pertencia, formando as entradas e saídas do banco de dados. 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.

Com isso, as entradas foram geradas para uma RNA a partir da FFT de cada amostra, realizando uma média das magnitudes da FFT para o agrupamento de diferentes faixas, no espectro de 50 Hz a 300 Hz, que é o intervalo onde o espectro de frequência da zabumba têm suas principais componentes. Estas entradas foram utilizadas para treinar uma rede perceptron multicamadas (PMC) com uma camada oculta utilizando Levenberg–Marquardt como algoritmo de otimização, tendo como métrica de erro o erro quadrático médio (EQM), sendo que posteriormente a avaliação do método é feito pelo Erro Percentual Absoluto Médio (EPAM), pois não necessariamente uma rede com o mínimo EQM terá também o mínimo EPAM (Ferreira-Paiva et al., 2022). Os passos realizados podem ser visualizados na Fig. 1

Em Velázquez Medina et al. (2019), a eficiência de uma rede PMC usada para estimar a potência de saída de parques eólicos têm sua eficiência melhorada ao adicionar mais informações de entrada à rede. Embora não se trate de processamento de áudio, esta rede se assemelha a que foi utilizada nesse trabalho por ter também apenas uma camada oculta e uma saída com valor escalar.

3. MATERIAIS E MÉTODOS

Nesta seção está presente a composição do banco de dados utilizado para a estimação da duração do passo base, com as variações com ruído e sem ruído. Também consta a seguir como as amostras do banco de dados foram processadas e suas variações para a aplicação em uma rede neural artificial, que também tem sua topologia descrita a seguir. Por fim, são apresentadas as métricas de avaliação dos resultados da rede.

3.1 Banco de dados

Os bancos de músicas utilizados para a realização deste projeto são os mesmos utilizados em Ferreira-Paiva et al. (2022) sem adição do ruído branco, com seus respectivos nomes e artistas ou bandas responsáveis pelas gravações das versões utilizadas. No total, foram utilizados 119 arquivos de músicas, sendo que 82 deles fazem parte do banco de dados sem ruído e 37 foram gravados em um ambiente de dança de forró, que será denominado como ruído real.

Após as amostras estarem preparadas com os pares entrada e saída, a FFT dos segmentos e as anotações de duração do compasso da música, o banco é dividido para que seja feita uma validação cruzada K-fold, sendo o valor de K = 10 *folds*. Com isso, reduz-se o impacto de uma seleção de um banco de dados tendencioso para a rede.

3.2 Geração de entradas

O impacto que o número de entradas tem no desempenho da rede é avaliado utilizando diferentes números de entradas para o treinamento do modelo. Os valores escolhidos foram 25, 50, 100 e 250 entradas, com base nos valores resultantes da resolução em Hz das entradas, onde os agrupamentos de faixas de frequência da FFT o espectro de 50 Hz a 300 Hz são 10 Hz, 5 Hz, 2,5 Hz e 1 Hz, respectivamente.

Também será avaliada a influência da faixa de frequência extraída dos segmentos de músicas, fixando o número de entradas a partir do melhor resultado encontrado para o número de entradas. As faixas escolhidas para este teste são os intervalos: [50; 300] Hz, com base no espectro da zabumba (Paiva et al., 2020; Ferreira-Paiva et al., 2022); (0; 300] Hz; (0; 800] Hz e, (0; 1600] Hz, onde o parênteses representa que o valor da extremidade não está presente e o colchetes denota que a extremidade está presente.

3.3 Rede perceptron multicamadas

A topologia de rede neural perceptron multicamadas (PMC) foi utilizada neste trabalho devido sua característica de aproximação universal de funções (Silva et al., 2016). Outra característica da PMC 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).

O modelo utilizado para o treinamento da rede é semelhante ao utilizado por Ferreira-Paiva et al. (2022), que contém:

- Camada de entrada com normalização z-score, que mantém a média igual a 0 e desvio padrão igual a 1, sendo representado por
$$\mathbf{Z} = \frac{\mathbf{X} - \mu}{\sigma}, \quad (2)$$
onde $\mathbf{X}$ é o lote de treinamento, μ e σ são a média e o desvio padrão de $\mathbf{X}$, respectivamente;
- Camada oculta, com 5 valores de neurônios testados segundo o Critério de Fletcher-Gloss descrito por
$$2\sqrt{n_0} + n_2 \leq n_1 \leq 2n_0 + 1, \quad (3)$$
onde n_0, n_1 e n_2 representam, respectivamente, as camadas de entrada, oculta e saída da rede (Silva et al., 2016), tendo como função de ativação a função tangente hiperbólica;
- Uma camada de saída com ativação linear e normalização z-score; e,
- Treinamento com métrica de perda Erro Percentual Absoluto Médio, EPAM, definido por
$$\text{EPAM}(\%) = \frac{1}{N}\sum_{i=1}^{N}\frac{|Y_i - Yd_i\,|}{Yd_i} \times 100\%, \quad (4)$$
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, com algoritmo de otimização ADAM e lote com tamanho 20, o que representa que os pesos da rede serão recalculados a cada 20 entradas.

Para evitar que o modelo decore o conjunto de treinamento (fenômeno conhecido como *overfitting*), o treinamento de cada rede foi finalizado sempre que EPAM dos dados de validação pararem de diminuir. Foi considerada a melhor rede aquela que obteve menor EPAM de todos os treinamentos realizados para cada método de extração de características.

Para avaliar a sensibilidade do modelo quando sujeito à amostras de natureza de ruído diferente do treinamento (com ruído real ou sem ruído), foi feito o cruzamento do banco de dados e então calculado o coeficiente de correlação de Pearson, que pode ser representado por
$$\rho = \frac{\text{cov}(\mathbf{X}, \mathbf{Y})}{\sqrt{\text{var}(\mathbf{X}) \cdot \text{var}(\mathbf{Y})}}, \quad (5)$$
na qual ρ representa o coeficiente de correlação, $\mathbf{X}$ os valores de saídas calculadas e $\mathbf{Y}$ os valores esperados. O valor de ρ pode variar de -1 à 1, sendo -1 representando uma correlação perfeita negativa, 0 sem correlação e 1 uma correlação perfeita positiva. Todo o processo descrito acima pode ser visualizado no fluxograma presente na Fig. 2. Por fim, os EPAMs dos melhores modelos para cada configuração de entradas são reunidos comparados com resultados da literatura encontrada.

4. RESULTADOS E DISCUSSÃO

As formas como o aumento do número de entradas do modelo e o aumento do espectro da FFT dos segmentos utilizados como amostras de entrada da rede neural serão analisadas por meio do desempenho para um banco de dados de teste do modelo. É importante destacar que o

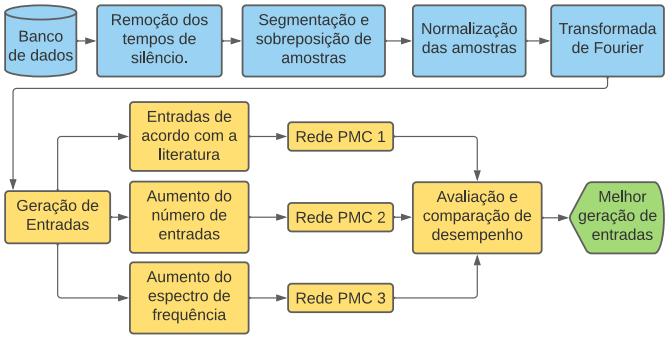


Figura 2. Fluxograma do processo de obtenção e comparação das redes. Onde foi feita a replicação de acordo com a literatura para padronizar as métricas de avaliação. Também aumentou-se o número de entradas e o espectro de frequência foi aumentado.

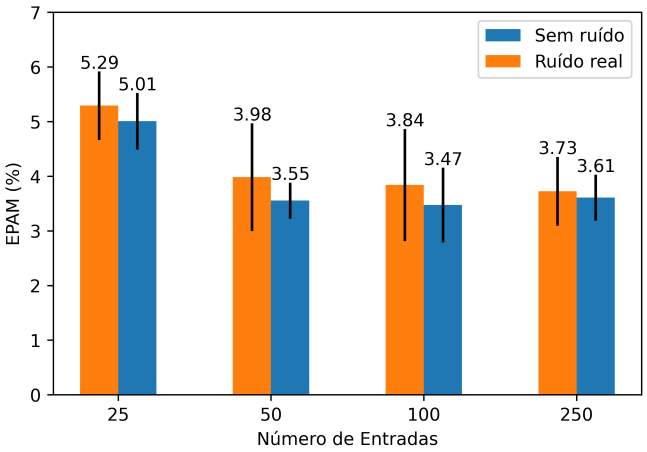
aumento do número de entradas do modelo aumenta a resolução das frequências de entrada, enquanto o aumento do espectro reduz a resolução e aumenta a variação de frequências de entrada.

4.1 Aumento de Entradas

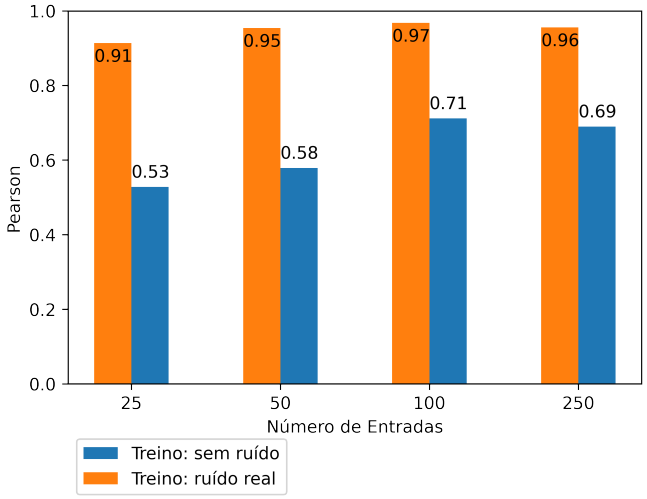
A Figura 3 apresenta as métricas de avaliação do desempenho. Na Fig. 3(a) são apresentados os resultados para o índice EPAM, dado por (4), e nas Figs. 3(b) e 3(c) para o coeficiente de correlação de Pearson, dado por (5), para os valores de 25, 50, 100 e 250 entradas. A Fig. 3(a) apresenta os valores do EPAM para a mesma natureza de ruído para a qual a rede foi treinada, enquanto as Figs. 3(b) e 3(c) apresentam o coeficiente de correlação de Pearson tanto para a mesma natureza de ruído quanto para a outra em relação ao treinamento da rede.

A Fig. 3(a) apresenta as menores médias e desvio padrão do EPAM dos K-folds para o banco de dados de ruído real e sem ruído para cada número de entradas. As linhas em preto representam o desvio padrão entre os K-folds e o valor representa a menor média entre os K-folds de acordo com a respectiva configuração de rede. Na mesma pode-se observar que o aumento de entradas para 50 reduziu o EPAM da rede de 5,29% para 3,98% e de 5,01% para 3,55% dos bancos de dados com ruído real e sem ruído, respectivamente. Entretanto, ao aumentar ainda mais o número de entradas da rede, o EPAM permanece próximo ao de 50 entradas, sendo que durante o treinamento, em alguns momentos, a rede não convergia devido ao alto número de parâmetros. Sendo assim, o valor de 50 entradas foi considerado como melhor do que os valores de 100 e 250 entradas.

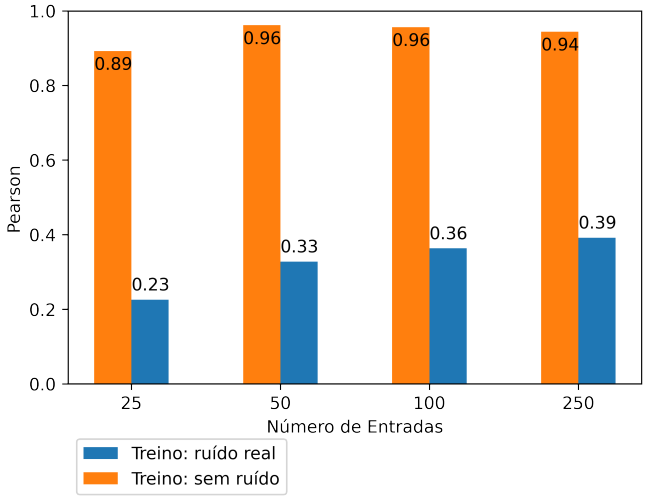
Ao calcular o EPAM para o cruzamento do banco de dados, notou-se que o mesmo não representava o quão distintos as amostras previstas estavam do desejado, como mostra a Fig. 4(b), onde consta as previsões da rede treinada sem ruído quando exposta à amostras com ruído real. O erro para este caso foi de 8,51%. Entretanto, como nota-se na Fig. 3(b), o coeficiente de correlação foi de apenas 0,58 para este caso, sendo o erro uma informação insuficiente para verificar a qualidade do modelo. Já na Fig. 4(a) são apresentadas as previsões da rede treinada com ruído real prevendo o mesmo conjunto de dados que



(a) EPAM das amostras de teste.

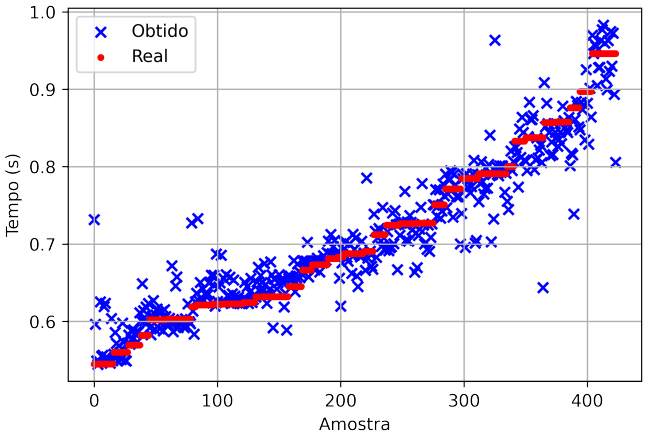


(b) Coeficientes de Pearson das amostras de teste com ruído real.

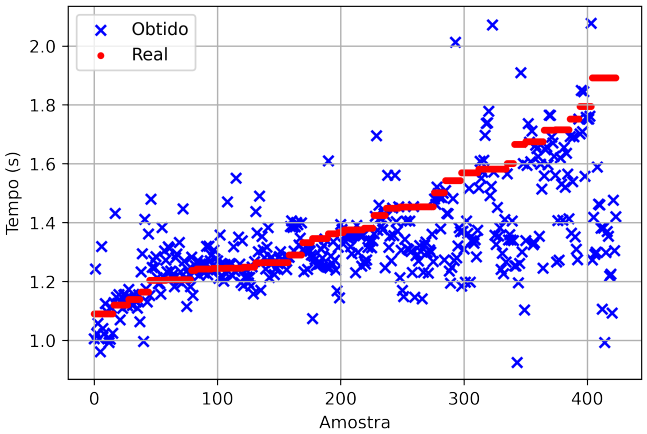


(c) Coeficientes de Pearson das amostras de teste sem ruído.

Figura 3. EPAM e Coeficientes de Pearson para cada número de entradas testado. Nota-se a redução do erro foi maior para o primeiro aumento do número de entradas que para os aumentos subsequentes. Entretanto, nenhum dos aumentos de entradas foi capaz de estimar amostras fora do ambiente de ruído para o qual a rede foi treinada.



(a) Rede treinada com ruído real.



(b) Rede treinada sem ruído.

Figura 4. Saídas estimadas e desejadas de amostras com ruído real das redes com 50 entradas. Nota-se que para a rede treinada com ruído real (a) os valores obtidos estão mais próximos que para a rede treinada sem ruído (b). Sendo que o comportamento de (b) os valores obtidos estão mais próximos de um valor constante quando comparado às previsões de (a)

foi utilizado para teste, onde o erro foi de 3,01% e coeficiente de correlação de 0,95, mostrando um comportamento bem mais similar entre os valores desejados e encontrados. Também é válido ressaltar que apenas o coeficiente de correlação não é suficiente para comprovar a eficácia da estimação, visto que o mesmo não é capaz de detectar a diferença entre dois conjuntos proporcionais.

Para o banco de dados com ruído real, ao calcular o coeficiente de correlação de Pearson para avaliar se a saída da rede tem relação com os valores reais, obteve-se os valores encontrados na Fig. 3(b), onde nota-se que as amostras de teste para a rede com menor EPAM não foi capaz de obter uma correlação relevante para o cruzamento do banco de dados (representado em azul). Porém, obteve-se uma correlação expressiva para a rede testada com o mesmo banco de dados que foi treinada (representado em laranja), indicando que todos os números de entradas testados nesse modelo não foram capazes de prever uma amostra fora do ambiente para o qual a rede foi treinada.

Já para o banco de dados sem ruído, obteve-se os valores encontrados na Fig. 3(c), na qual os resultados foram semelhantes às previsões com ruído real. Entretanto, o coeficiente de correlação do cruzamento do banco de dados se mostrou menor quando comparados ao cruzamento do banco de dados da Fig. 3(b). Isto se deve às músicas presentes no banco de dados com ruído real também estarem presentes no banco de dados sem ruído. Portanto, a rede treinada sem ruído previu no teste somente músicas com as quais teve contato no treinamento, porém, acrescidas de ruído, o que não acontece no cruzamento reverso.

4.2 Aumento do Espectro de Frequência

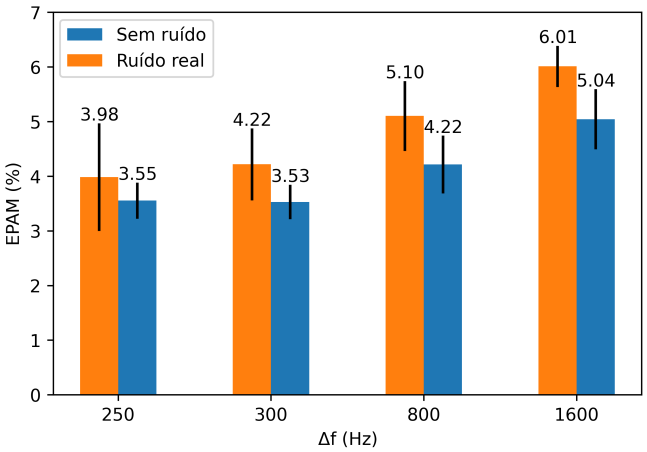
A Figura 5 apresenta as métricas de avaliação do desempenho. Na Fig. 5(a) são apresentados os resultados para o índice EPAM, dado por (4), e nas Figs. 5(b) e 5(c) para o coeficiente de correlação de Pearson, dado por (5), para as faixas de [50; 300] Hz, (0; 300] Hz, (0; 800] Hz e (0; 1600] Hz; que estão representados pela variação de frequência de $\Delta_f = 250$, $\Delta_f = 300$, $\Delta_f = 800$, $\Delta_f = 1600$, respectivamente, utilizando o número de entradas igual a 50. A Fig. 5(a) apresenta os valores do EPAM para a mesma natureza de ruído para a qual a rede foi treinada, enquanto as Figs. 5(b) e 5(c) apresentam o coeficiente de correlação de Pearson tanto para a mesma natureza de ruído quanto para a outra em relação ao treinamento da rede.

A Figura 5(a) apresenta as menores médias e desvio padrão do EPAM dos K-folds para o banco de dados de ruído real e sem ruído para cada variação de frequência. Nota-se que o EPAM cresce para faixas de frequência maiores, indicando que as magnitudes das frequências acima de 300 Hz não contribuem positivamente para a eficácia da rede.

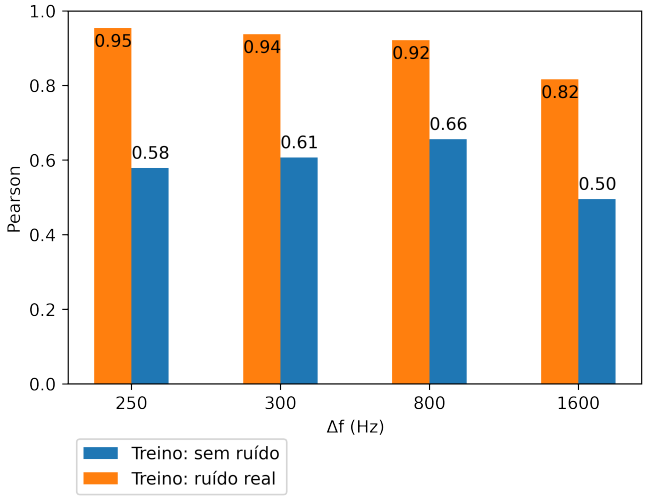
Nas Figuras 5(b) e 5(c), nota-se comportamentos semelhantes aos das correlações das Figs. 3(b) e 3(c) para o cruzamento do banco de dados, onde as previsões das amostras com ruído real são ligeiramente maiores. Entretanto, ainda são insuficientes para afirmar que a rede é capaz de prever amostras de natureza de ruído diferente. Quando as amostras são submetidas a uma rede treinada com amostras no mesmo ambiente de ruído, os coeficientes de correlação resultaram em valores mais próximos a 1 pra todos os casos exceto para a variação de frequência de 1600 Hz com ruído real. Logo, mesmo que o EPAM para este caso seja próximo a 6%, esta configuração de entrada não é ideal para a utilização pois a saída da rede não tem relação com o desejado.

A provável razão do aumento do EPAM para os valores mais altos de Δ_f pode ser observada na Fig. 6, onde nota-se maior variação das magnitudes das frequências de 600 Hz a 1600 Hz ao longo do tempo. Este fenômeno resulta em diferentes valores de entrada para uma mesma saída da RNA, dificultando o treinamento da rede por ter maiores variações de entrada para uma mesma saída.

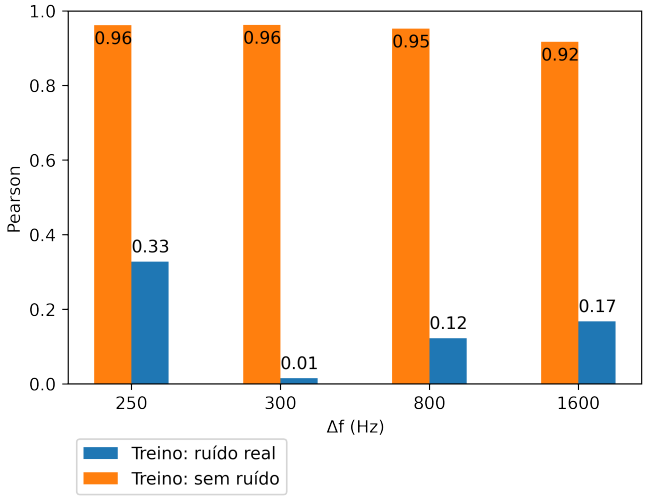
Embora o número de parâmetros das redes tenha crescido ao aumentar o número de entradas de 25 para 50 (de 730 para 2549 para músicas com ruído real e de 1378 para 5201 para músicas sem ruído), o tempo de treinamento para a rede foi mais afetado pelo tamanho



(a) EPAM das amostras de teste.



(b) Coeficientes de Pearson das amostras de teste com ruído real.



(c) Coeficientes de Pearson das amostras de teste sem ruído.

Figura 5. EPAM e Coeficientes de Pearson para cada faixa de frequência. Percebe-se que ao aumentar o espectro de frequência para além de 300 Hz também resultou no aumento do erro, indicando que o espectro de interesse está nas frequências mais baixas. Esta alteração também não auxiliou na previsão de amostras fora do ambiente de ruído para o qual a rede foi treinada.

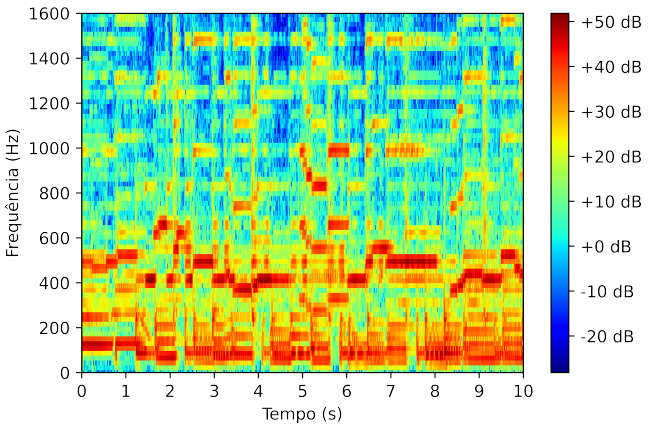


Figura 6. Espectrograma de uma das músicas utilizadas. Percebe-se que as maiores variações das intensidades estão nas faixa de frequência mais altas, o que aumenta o número de entradas diferentes para uma mesma saída. Com diferentes entradas para uma mesma saída, a função para a estimação da duração do passo base será mais complexa.

do banco de dados, Sendo o tempo de treinamento para músicas com ruído real de aproximadamente 20 ms e para músicas sem ruído de 30 ms para cada variação de neurônios na camada oculta e de entrada testada. Já o tempo de inferência foi menor que 1 μs por amostra para qualquer variação de rede, ressaltando que em uma aplicação em tempo real o tempo para a extração da amostra pelo microfone será de aproximadamente 3 s, ou seja, mesmo com o aumento do tempo de treinamento e inferência da rede o mesmo não representará a maior parte do tempo total de processamento do sinal até a estimação do compasso da música. Logo, foi necessário utilizar uma rede com mais parâmetros para melhorar a estimação da rede em cerca de 1% em termos absolutos. No entanto, o tempo adicional não é relevante considerando o tempo de extração da amostra.

Em resumo, o aumento do número de entradas da rede para a mesma faixa de frequência foi capaz de reduzir o erro em 24,76% e 29,14% do EPAM para músicas com ruído real e sem ruído, respectivamente. Com isso, aumentar o número de entradas para 50, mantendo a faixa de frequência de 50 Hz à 300 Hz foi o método de extração de características que proporcionou o menor erro. Vale ressaltar que o aumento de entradas para 100 e 250 resultou em erros menores, mas, durante o treinamento, houveram casos em que a rede não convergiu devido à quantidade elevada de parâmetros a serem ajustados em proporção ao banco de dados. Além disso, nenhum método foi capaz de prever corretamente a duração do compasso quando submetido a uma amostra de natureza de ruído diferente ao utilizado durante o treinamento da rede.

4.3 Comparação com a literatura

Levanto em consideração alguns trabalhos encontrados na literatura (Paiva et al., 2020; Ferreira-Paiva et al., 2022), os principais indicadores de desempenho foram compilados na Tabela 1, assim como o menor EPAM das redes treinadas para cada variação de entradas. Na Tabela 1, nota-se que as alterações feitas na rede de 25 entradas reduziu o erro da melhor rede de 5,54% para 4,51% da rede treinada com ruído real e de 5,06% para 4,21% da rede treinada com amostras sem ruído, comprovando que, de fato, o treinamento da rede com EQM não encontrou a rede com menor EPAM no teste. Também observa-se a redução do erro da melhor rede para quase 3% em ambos os bancos de dados ao utilizar 50 entradas, lembrando que o desvio padrão dos K-folds para o banco de dados sem ruído foi menor, indicando uma melhor confiabilidade desse banco de dados possivelmente pelo mesmo conter mais amostras. Por fim, embora os erros para os cruzamentos dos bancos de dados esteja menor que a literatura em diversas configurações de entrada, nenhuma foi considerada solução para o problema de amostras de natureza diferente por causa do coeficiente de correlação ser baixo.

Tabela 1. EPAM das amostras de teste da rede com menor erro para o mesmo banco de dados para a qual a rede foi treinada (Direto) e para o banco de dados de natureza de ruído diferente (Cruzamento). *Modelo com 25 entradas e com $\Delta_f = 250$. **Valor aproximado do gráfico.

	EPAM (%)			
Treinamento	Ruído Real		Sem Ruído	
Entradas	Direto	Cruzamento	Direto	Cruzamento
ent = 25	4,51	17,11	4,21	9,84
ent = 50	3,01	13,24	3,02	8,51
ent = 100	2,71	11,87	2,58	7,60
ent = 250	2,70	12,94	2,95	8,16
$\Delta_f = 300$	3,32	18,65	2,90	8,98
$\Delta_f = 800$	4,02	18,36	3,52	8,32
$\Delta_f = 1600$	5,44	13,54	4,04	10,29
Paiva et al. (2020)*	-	-	3,41	-
Ferreira-Paiva et al. (2022)*	5,54	-	5,06	$\approx 12,8$**

5. CONCLUSÃO

Este trabalho analisou duas alterações nas extrações de características de músicas de forró existente na literatura. O aumento do número de entradas da rede para 50 entradas, mantendo a faixa de frequência de 50 Hz a 300 Hz, foi capaz de reduzir o EPAM em mais de 20% para ambos os bancos de dados sem ruído e com ruído real. Porém, mesmo esta configuração de entradas não fez com que a rede fosse capaz de prever amostras fora do ambiente de ruído para a qual foi treinada.

Ainda, se faz necessário um estudo do quanto a mistura dos bancos de dados irá beneficiar a previsão de diferentes ambientes, levando em conta diferentes proporções de músicas em diferentes ambientes de ruído, que será o próximo objetivo de estudo em trabalhos futuros. Dessa forma, espera-se que haja avanços para a criação de um aplicativo que transmita o ritmo da música, por meio de estímulos táteis ou visuais para surdos, para ajudar no aprendizado da dança pessoas surdas ou com alguma deficiência auditiva, melhorando sua qualidade de vida por meio de mais interações sociais.

REFERÊNCIAS

Bossey, A. (2020). 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*, 11(1), 6–25. doi:10.1108/IJEFM-03-2019-0022. URL https://www.emerald.com/insight/content/doi/10.1108/IJEFM-03-2019-0022/full/html.

Brasil (2002). Lei nº 10.436, de 24 de abril de 2002. *Diário Oficial [da] República Federativa do Brasil*. URL http://www.planalto.gov.br/ccivil_03/leis/2002/l10436.htm.

Brasil (2005). Decreto nº 5.626, de 22 de dezembro de 2005. *Diário Oficial [da] República Federativa do Brasil*. URL http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2005/Decreto/D5626.htm.

Cai, L. and Cai, Q. (2019). Music creation and emotional recognition using neural network analysis. *Journal of Ambient Intelligence and Humanized Computing*, (0123456789). doi:10.1007/s12652-019-01614-6. URL https://doi.org/10.1007/s12652-019-01614-6.

Chaveiro, N., Duarte, S., Freitas, A., Barbosa, M., Porto, C., and Fleck, M. (2014). Qualidade de vida dos surdos que se comunicam pela língua de sinais: revisão integrativa. *Interface - Comunicação, Saúde, Educação*, 18, 101–114. doi:10.1590/1807-57622014.0510.

de Lacerda, C.B.F. (2006). A inclusão escolar de alunos surdos: o que dizem alunos, professores e intérpretes sobre esta experiência. *Cadernos CEDES*, 26(69), 163–184. doi:10.1590/s0101-32622006000200004.

de Paula, T.R.M. and Pederiva, P.L.M. (2017). Musical Experience in Deaf Culture. *International Journal of Technology and Inclusive Education*. doi:10.20533/ijtie.2047.0533.2017.0136. URL https://www.semanticscholar.org/paper/Musical-Experience-in-Deaf-Culture-Paula-Pede\riva/1e450c197d715f94ee8b062a5eb85f26c33adf5e.

de Quadros Junior, A.C. and Volp, C.M. (2005). Forró universitário: a tradução do forró nordestino no sudeste brasileiro. *Motriz. Journal of Physical Education. UNESP*, 117–120.

Ferreira-Paiva, L., Lopes, H.G., Alfaro-Espinoza, E.R., Felix, L.B., and Neves, R.V.A. (2022). Towards a device for helping deaf people to dance: estimation of forro bar length using artificial neural network. *IEEE Latin America Transactions*, 20(6), 970–976. doi:10.1109/tla.2022.9757740. URL https://doi.org/10.1109\%2Ftla.2022.9757740.

Fox, M.L., James, T.G., and Barnett, S.L. (2020). Suicidal behaviors and help-seeking attitudes among deaf and hard-of-hearing college students. *Suicide and Life-Threatening Behavior*, 50(2), 387–396. doi:https://doi.org/10.1111/sltb.12595. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/sltb.12595.

FREITAS, F.B. (2019). Benefícios psicológicos da prática da dança em pessoas com diagnóstico de ansiedade e depressão: Revisão bibliográfica. Monografia (Bacharel em Psicologia), UFM (Universidade Federal do Maranhão), São Luís, Maranhão, Brasil.

Haguiara-Cervellini, N. (2003). *A musicalidade do surdo: representação e estigma*. Plexus Editora.

IBGE (2012). Censo demográfico : 2010 : características gerais da população, religião e pessoas com deficiência. *IBGE*. URL https://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd_2010_religiao_deficiencia.pdf.

Mirzaei, M., Kán, P., and Kaufmann, H. (2020). EarVR: Using Ear Haptics in Virtual Reality for Deaf and Hard-of-Hearing People. *IEEE Transactions on Visualization and Computer Graphics*, 26(5), 2084–2093. doi:10.1109/TVCG.2020.2973441.

Paiva, L.F., Lopes, H.G., Felix, L.B., and Neves, R.V.A. (2020). Estimação do compasso musical do forró utilizando rede perceptron multicamadas. In *Anais do XXIII Congresso Brasileiro de Automática*. Porto Alegre.

Salamon, J. and Bello, J.P. (2017). Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *IEEE Signal Processing Letters*, 24(3), 279–283. doi:10.1109/LSP.2017.2657381.

Santos, A., Tang, L., Loke, L., and Martinez-Maldonado, R. (2018). You are off the beat!: Is accelerometer data enough for measuring dance rhythm? 1–8. doi:10.1145/3212721.3212724.

Santos, T.M., Vieira, L.C., and Faria, C.A. (2013). Deficiência auditiva e mercado de trabalho: uma visão de empregadores da cidade de Uberlândia-MG. *Psicologia: teoria e prática*, 15, 92 – 103. URL http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-36872013000200007&nrm=iso.

Schoenberg, A. (1990). *Fundamentos da composição musical*. Edusp.

Sharp, A., Bacon, B.A., and Champoux, F. (2020). Enhanced tactile identification of musical emotion in the deaf. *Experimental Brain Research*, 238(5), 1229–1236. doi:10.1007/s00221-020-05789-9. URL https://doi.org/10.1007/s00221-020-05789-9.

Sharp, A., Houde, M.S., Bacon, B.A., and Champoux, F. (2019). Musicians show better auditory and tactile identification of emotions in music. *Frontiers in Psychology*, 10(AUG), 1–7. doi:10.3389/fpsyg.2019.01976.

Silva, I.N., Spatti, D.H., and Flauzino, R.A. (2016). *Redes Neurais Artificiais para Engenharia e Ciências Aplicadas: Fundamentos Teóricos e Aspectos Práticos*. Artliber, São Paulo, 2ª edição edition.

Solanki, A. and Pandey, S. (2019). Music instrument recognition using deep convolutional neural networks. *International Journal of Information Technology*. doi:10.1007/s41870-019-00285-y. URL https://doi.org/10.1007/s41870-019-00285-y.

Velázquez Medina, S., Carta, J.A., and Portero Ajenjo, U. (2019). Performance sensitivity of a wind farm power curve model to different signals of the input layer of anns: Case studies in the canary islands. *Complexity*, 2019, 2869149. doi:10.1155/2019/2869149. URL https://doi.org/10.1155/2019/2869149.

Yu, D., Duan, H., Fang, J., and Zeng, B. (2020). Predominant Instrument Recognition Based on Deep Neural Network with Auxiliary Classification. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 28, 852–861. doi:10.1109/TASLP.2020.2971419. URL https://ieeexplore.ieee.org/document/8979336/.