Towards a Device for Helping Deaf People to Dance: Estimation of "Forró" Bar Length using Artificial Neural Network

Lucas Ferreira-Paiva, Hugo G. Lopes, Elizabeth R. Alfaro-Espinoza, Leonardo B. Felix, and Rodolpho V. A. Neves, *Member, IEEE*

***Abstract*—Dance has the potential to improve people's quality of life, as well as assist to decrease depression and anxiety. However, the lack of technologies capable of exploring alternative senses of hearing limits music and dance's beneficial effects on listeners. In order to find a model capable of being implemented in accessible devices, this work evaluated the use of a model based on neural networks to estimate the *forró* music bar length. Model variations were trained for seven datasets composed of mixes of music samples without noise, with real noise and with white noise. For each dataset, the best variation was selected and these were evaluated for the same real noise samples. The model variations that were presented to samples with real noise in the training estimated the bar duration with an average percentage error of less than 7% in the test step, being significantly better the model trained only with real sample. The evaluated model was able to estimate the length of the *forró* music bar length, even in real scenarios, as long as it was presented in this scenario during training. Increased database diversity and the use of data augmentation techniques can lead to improvements in the generalizability of the model. The simplicity of the evaluated model and its ability to learn when properly trained, indicate its potential to be used, in real time, on a mobile device to pass the rhythm of *forró* music to deaf and hard of hearing (D/HH) people.**

***Index Terms*—Rehabilitation engineering, music, inclusion, multilayer perceptron, Brazil.**

I. Introdução

A perda de audição afeta cerca de 430 milhões de pessoas no mundo e tende a atingir 700 milhões de pessoas até 2050 [1]. Até meados do Séc. XX, a história das pessoas surdas e com deficiência auditiva (S/DA) foi marcada pela imposição do oralismo e proibição da comunicação gestual [2], [3]. Atualmente, a comunicação gestual é reconhecida como Língua de Sinais (LS) [4]. Dentro da população S/DA, pode se caracterizar como "surdos" os indivíduos que se reconhecem como parte da "Cultura Surda", logo falantes da LS [5].

Para muitos surdos, a comunicação por LS é preferível à recuperação da audição por aparelhos auditivos e implantes cocleares [6]. A dificuldade de comunicação entre as pessoas S/DA com seus pais e cuidadores, precariza situações do cotidiano, como a alimentação [7] e resulta em maior frequência de doenças mentais e intensão suicida que pessoas ouvintes [8], [9]. Entender as demandas das pessoas com deficiências faz parte dos desafios da engenharia de reabilitação [10] e a dança pode ser um dos caminhos devido seu potencial de inclusão [11].

Um estudo qualitativo com crianças mostrou que a dança possibilita ver a diferença como algo comum e a apreciar a diversidade [12]. Na mesma linha, meta análises sobre o efeito da dança e dançaterapia na saúde mostraram que a dança pode produzir aumento na qualidade de vida, bem-estar, humor, afeto, imagem corporal e resultados clínicos, além de diminuição da depressão e ansiedade [13], [14]. Ainda, a participação de crianças surdas em aulas de dança com o uso de fones vibratórios proporciona melhora da autoestima [15].

A percepção da música não se restringe à modalidade auditiva, uma vez que, as baixas frequências da música geram vibrações que podem ser sentidas no corpo ou através de objetos [16]. Devido a essa possibilidade, a pelo menos 30 anos já se estuda estratégias para passar o ritmo das músicas por vibração [17]. Pode-se destacar o trabalho de [18] que propôs um sistema de detecção de batida para auxiliar dançarinos de Salsa com comandos de voz/vibração e o trabalho de [19] apresentou um protótipo para ajudar pessoas surdas a sentirem o ritmo da música através de luz e vibração.

Pessoas S/DA são igualmente capazes de sincronizar com o ritmo da música através de estímulos táteis [20] e podem detectar emoções em músicas através de vibração com desempenho superior a ouvintes [21]. Em contra partida, um estudo recente identificou 83 instrumentos musicais digitais inclusivos e, destes, apenas 5 eram destinados a pessoas S/DA [22], mostrando que a inserção de surdos em contextos de música precisa ser mais explorada.

No Brasil, o forró, uma festa que virou um estilo musical, é dançado por todas as camadas da sociedade. Os passos básicos do forró são realizado ao longo de dois compassos [23], portanto, essa componente permite sinalizar a distancia temporal entre o início e o fim de um passo, indicando a velocidade para se dançar a música. Devido essa relação, a duração do compasso é utilizada em aplicações que almejam obter o ritmo de músicas de forró [24].

Visando a inclusão do público surdo em atividades culturais envolvendo o forró, um trabalho prévio do grupo [25] propôs

Lucas Ferreira-Paiva faz parte do Programa de Pós-Graduação em Ciência da Computação, Departamento de Informática, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil, e-mail:lucas.f.paiva@ufv.br.

Hugo G. Lopes, Rodolpho V. A. Neves e Leonardo B. Felix fazem parte do Departamento de Engenharia Elétrica, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil, e-mail:{hugo.lopes, leobonato, rodolpho.neves}@ufv.br.

Elizabeth R. Alfaro-Espinoza faz parte do Programa de Pós-Graduação em Bioinformática, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil, e-mail:elizaespinoza@ufmg.br.

um modelo computacional que estima a duração do compasso de musicas de forró utilizando uma rede perceptron multicamadas (PMC). No entanto, para que o modelo não tenha queda brusca de desempenho, antes de ser usado para passar o ritmo de musicas de forró para surdos por estímulos táteis, é necessário que seja treinado e validado para um cenário com ruídos reais [26].

Não foram encontrados trabalhos além de [25] que estimaram a duração do compasso diretamente, somente do BPM [27]–[31], da estrutura de divisão do compasso [32] e de ambas métricas simultaneamente [33]. Essas métricas até podem ser combinadas para se obter a duração de um compasso [24], mas exigiria que ambas fossem estimadas corretamente.

Portanto, o objetivo deste trabalho é ajustar e avaliar o modelo de [25], para músicas com ruídos reais. As contribuições em relação ao trabalho anterior são: (i) a ampliação do banco de dados proposto, com o acréscimo de 42 músicas, no qual foi levantado as respectivas durações dos compassos; (ii) a criação de um banco de dados com músicas de forró com ruídos de um espaço de dança; (iii) a avaliação do uso de ruído branco para simular o ruído real; e (iv) o ajuste e avaliação do modelo proposto por [25] para um cenário real.

II. Criação do Banco de Dados

Com a parceria de uma instrutora de forró e um projeto presente no campus de Viçosa-MG da Universidade Federal de Viçosa foram selecionadas músicas populares nos eventos de forró no campus. Foram acrescentadas 42 músicas ao banco de [25], totalizando 82 músicas, majoritariamente de Forró Pé-de-serra e Forró Universitário, segundo caracterização de [34], com diversidade rítmica, de músicas "mais lentas" às músicas "mais rápidas". Os títulos das músicas selecionadas, os respectivos interpretes da versão escolhida e a duração do compasso medida para cada música podem ser observados no repositório GitHub disponível em https://github.com/NIASUFV/ForAll.

A. Geração dos Datasets

Na Fig. 1 é apresentado um fluxograma com o processo de divisão do banco de dados para a composição dos *datasets* de músicas sem ruídos, com ruídos reais e com ruído branco. As letras S, R e B, foram utilizadas para indicar Sem Ruído, Ruído Real e Ruído Branco, respectivamente, conforme listado a seguir:

- **S** — Sem Ruído
- **R** — Ruído Real
- **B** — Ruído Branco
- **SR** — Sem Ruído + Ruído Real
- **SB** — Sem Ruído + Ruído Branco
- **RB** — Ruído Real + Ruído Branco
- **SRB** — Sem Ruído + Ruído Real + Ruído Branco

a) Músicas sem Ruído: O banco de dados sem ruído é composto pelos 82 arquivos de áudio das músicas selecionadas em MP3, a uma taxa de amostragem de 44,1 kHz.

b) Músicas com Ruído Real: O modelo avaliado será utilizado em um aplicativo móvel que ficará transmitindo o ritmo da música para os surdos por meio de estímulos táteis, seja pela vibração do celular ou por um dispositivo auxiliar

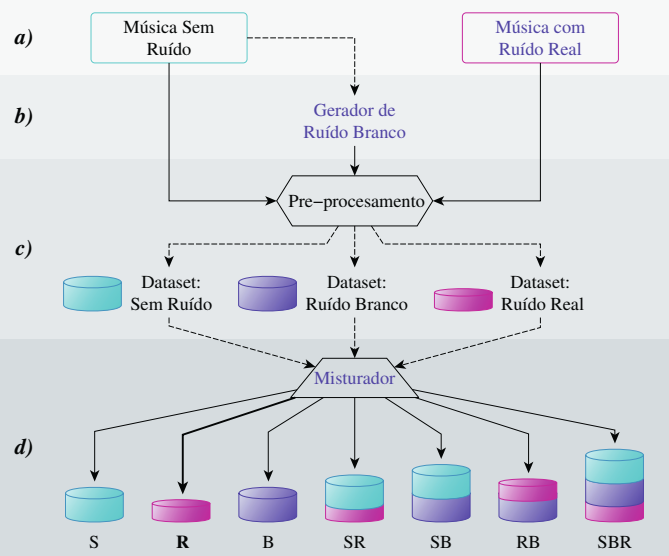


Fig. 1. Processo de criação dos *datasets* a partir das músicas sem ruído e músicas com ruído real de um espaço de dança.

de forma similar aos trabalhos de [18] e [35]. Portanto, se fez necessário um banco de dados real para minimizar a queda significativa do desempenho de modelos de recuperação musical criados em *Closed World* quando submetidos a dados de *Real World* [26].

De forma a obter um banco de dados com ruídos similares à aplicação, 37 faixas foram reproduzidas em um espaço de dança e regravadas por um celular Samsung Galaxy J5 localizado no bolso de um dançarino. As gravações foram feitas no formato WAV, a uma taxa de amostragem de 44,1 kHz, formando o banco "Músicas com Ruído Real".

c) Gerador de ruído branco: Foi adicionado ruído branco nas músicas sem ruído e o nível de ruído foi ajustado a um nível de relação sinal ruído de 30 dB. A inserção do ruído foi feita nas 82 músicas do banco de dados.

d) Preprocessamento: Todos os arquivos de áudio receberam o mesmo tratamento independentemente de ter ruído ou não. Todas as etapas estão enumeradas a seguir.

1) Recorte dos 20 segundos iniciais e finais para remover os períodos de silêncio da música;
2) Segmentação dos arquivos em trechos de três segundo com sobreposição de um segundo; e
3) Normalização de cada trecho dividindo-se pelo valor eficaz do trecho, para eliminação do efeito de volume.

e) Misturador: Partindo do pressuposto que apresentando amostras com e sem ruído durante o treinamento da rede neural a rede aprende a priorizar as entradas fundamentais, todas as combinações possíveis foram feitas com os três bancos de dados existentes. Os bancos de dados mistos foram criados concatenando as amostras de um, dois ou três *datasets* presentes.

B. Anotação dos Dados

A extração do tempo do compasso foi feita de forma indireta a partir do tempo gasto para a execução de um passo base frente e trás (PBFT). O PBFT pode ser descrito em oito

posições (P0-P7), como apresentado na Fig. 2, e é realizado ao longo de dois compassos sendo repetido indefinidamente [24].

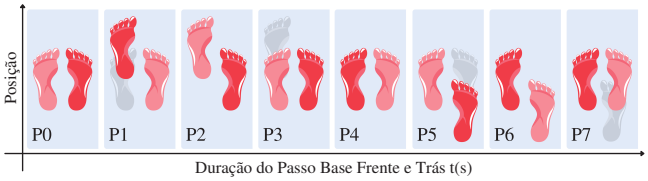


Fig. 2. Ilustração de um passo base completo segundo [24]. O vermelho em tom mais escuro indica o pé que está recebendo a maior parte do peso do corpo.

Para medir o tempo do compasso de cada música selecionada, uma instrutora de forró dançou todas as 82 músicas realizando somente o PBFT. A medição consistiu em cronometrar o tempo gasto para realizar um passo base completo, 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 PBFT por dois. Assumiu-se que a duração do compasso não variou ao longo da música, desta forma cada amostra está associada à duração do compasso da música que pertence.

III. Criação do Modelo de Estimação

O modelo proposto por [25], apresentado na Fig. 3, possui um etapa de extração de característica, por meio da Transformada Rápida de Fourrier, e uma etapa de estimação do compasso, por meio de uma rede PMC com uma camada oculta. O modelo foi treinado para cada um dos *datasets* e avaliado com validação *k-fold*, para definir o número de neurônios na camada oculta otimizado do modelo para cada *dataset*. Os melhores modelos foram avaliados com o *dataset* de ruído real (R), a fim de definir o *dataset* mais apropriado para capacitar o modelo a trabalhar em cenários reais.

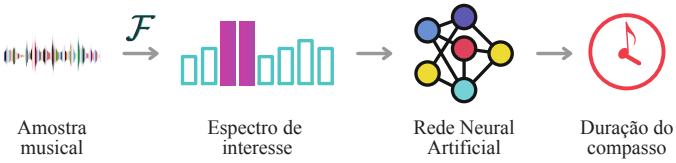


Fig. 3. Modelo para a estimação do ritmo de músicas de forró pelo compasso musical implementado por [25].

A. Características para Alimentação do Modelo

Cada amostra do banco de dado consiste em um par de entradas-saída. As entradas consistem em 3 s de música de forró. As características selecionadas são as principais componentes do espectro de frequência mais grave da zabumba (50 a 300 Hz), retiradas a partir do espectro completo de cada trecho segmentado da música. O espectro é dividido em 25 faixas com espaçamento de 10 Hz. Assim, cada entrada corresponde à média das amplitudes das componentes nas frequências correspondente à respectiva faixa (50-60, 60-70, ..., 290-300 Hz).

Quanto maior a amplitude da componente na frequência, maior a ocorrência do som caracterizado por esta frequência. Portanto, para uma amostra de 3 s, é esperado que músicas mais rápidas possuam maiores amplitudes das componentes na faixa da zabumba do que músicas mais lentas. Partindo destes efeitos, o modelo proposto por [25] mostrou-se capaz de estimar a duração do compasso para músicas de forró sem ruído.

B. Estrutura da Rede

A rede neural escolhida foi o perceptron multicamadas com uma única camada oculta, devido sua característica de aproximação universal de funções [36]. 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 [36]. A seguir estão listados os hiper-parâmetros para qual o modelo foi treinado.

- Número de entradas: 25;
- Número de saídas: 1;
- Número de neurônios na camada oculta: [11, 12, ..., 51];
- Função de ativação da camada oculta: Tangente hiperbólica;
- Função de ativação da camada de saída: Linear; e
- Taxa de aprendizagem: 0,001.

Todos os hiper-paramentos são fixos, exceto o número de neurônios na camada oculta que variou de 11 a 51 conforme o critério de Fletcher-Gloss [36].

C. Treinamento da PMC

Conforme apresentado na Fig. 1, foram criados sete *datasets* para treinar as redes: *dataset* sem ruído (S), com ruído real (R), com ruído branco (B) e com as combinação destes (SN, SB, RB e SRB). Para cada *dataset* foram treinadas 40 variações do modelo mudando o número de neurônios na camada oculta.

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 modelar essa relação de entrada e saída [36]. O método de Levenberg–Marquardt foi utilizado para otimizar o treinamento. O método *early stopping* foi implementado para interromper o treinamento precocemente e evitar *overfitting*. Para escolha do melhor modelo foi utilizada validação cruzada *k-fold* com k=7, onde 6 *folds* foram reservados para o treinamento, que equivale aproximadamente a 86% das amostras de cada *dataset* e um *fold* para teste. Para cada uma das sete interações da validação *k-fold*, foram feitos dez treinamentos sorteando novos pesos iniciais, abrindo oportunidade de condições iniciais mais favoráveis ao aprendizado da rede. Ao todo foram feitos 19.600 treinamentos.

D. Desempenho dos Modelos

A medida de desempenho utilizada em todos os testes foi o erro percentual médio (EPM) entre a resposta esperada e

a saída da rede. Também foi observado o desvio padrão DP do EPM encontrado para cada *fold* da validação *k-fold* para avaliar a variância do erro do modelo.

a) Amostras da mesma natureza do treino: Nesta etapa os modelos foram avaliados a partir dos desempenhos na fase de teste, que contou com amostras da mesma natureza do *dataset* de treino. Para todos os *dataset* avaliou-se o efeito do número de neurônios na camada oculta da rede PCM no desempenho no teste. O modelo com maior desempenho para cada *dataset* foi selecionado como candidato a melhor modelo para estimar a duração do compasso no cenário real.

b) Amostras com ruído real: Os modelos selecionados com melhor desempenho na fase de teste quando treinados e testados com os *datasets* S, B, SN, SB, RB e SRB, tiveram seus desempenhos avaliados quando submetidos às amostras de teste que foram utilizado para testar os modelos treinados com o *dataset* de ruído real.

Como todos os modelos foram avaliados com o mesmo conjunto de amostras nesta etapa, o teste *t student* foi utilizado comparando todos os modelos por pares a fim de avaliar a significância de eventuais diferenças encontradas. Este teste possibilita selecionar o melhor modelo para estimar a duração do compasso para amostras reais.

Nessa etapa, além de definir qual o melhor modelo para estimar o compasso em um cenário real, é possível observar o quanto cada modelo altera o desempenho quando apresentado a um banco de dados de natureza diferente da qual foi treinado. Permitindo avaliar a necessidade de se utilizar um *dataset* fiel ao cenário que o modelo será aplicado.

IV. Resultados

Os 82 arquivos de música sem ruído somados aos 37 arquivos de áudio gravados com ruído real resultaram em 9.632 amostras de três segundos. As durações dos compassos das músicas selecionadas variaram de 1,002 s até 1,92 s.

A. Desempenho de Teste para Cada dataset

Os desempenhos das redes na etapa de teste possibilitou avaliar o modelo proposto por [25] variando o número de neurônios na camada oculta, além de selecionar a configuração de rede mais adaptada ao problema para cada *dataset*.

Na Fig. 4, é possível observar que para todos os *datasets* houve queda acentuada do EPM entre 11 e 30 neurônios e a partir dessa faixa ocorre uma estabilização do erro, indicando que os benefícios de aumentar o número de neurônios na camada oculta ficam cada vez menores. Para uma implementação em dispositivo móvel ou embarcado, é mais vantajoso escolher uma rede com menor número de neurônios de camadas ocultas sem perda efetiva de desempenho.

O número de neurônios na camada oculta que ocasionou em melhor desempenho do modelo e os desempenhos para cada caso, destacado por uma estrela vermelha na Fig. 4, são apresentados na Tabela I.

Os resultados mostrados na Tabela I indicam que o modelo avaliado foi capaz de estimar a duração do compasso de músicas de forró com erro percentual médio EPM menor que 6%, sustentando a potencialidade do modelo proposto por

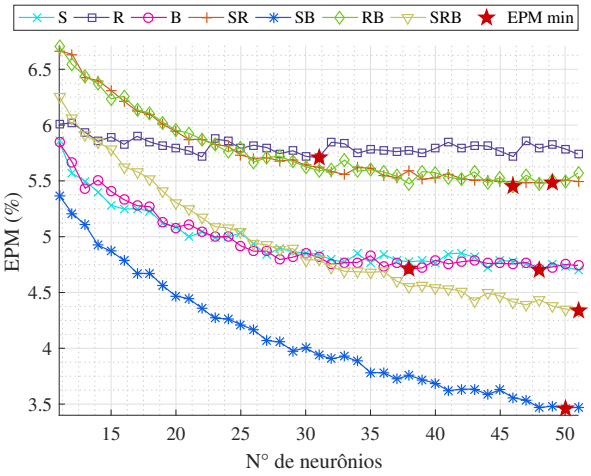


Fig. 4. Variação do EPM no teste para cada *dataset* em função do número de neurônios na camada oculta. A estrela vermelha indica a rede com menor EPM para cada *dataset*.

TABELA I
NÚMERO DE AMOSTRAS UTILIZADAS PARA O TESTE, NÚMERO DE NEURÔNIOS NA CAMADA OCULTA E DESEMPENHO DAS MELHORES REDES, PARA CADA *dataset* DE ACORDO COM O MENOR EPM E O DESVIO PADRÃO (DP) OBTIDOS COM A VALIDAÇÃO *k-fold*.

Dataset	Amostras	Neurônios	EPM(%)	DP(%)
S	921	48	5,059	0,047
R	458*	31	5,541	0,116
B	921	38	4,600	0,074
SR	1313	49	5,344	0,163
SB	1843	50	3,209	0,116
RB	1313	46	5,313	0,145
SRB	2235	51	4,176	0,180
[25]	425	87	3,408	-

*As amostras de teste do dataset R em destaque também foram utilizadas para avaliar o desempenho dos demais modelos em cenário real.

[25] para embarcar um aplicativo móvel que passe o ritmo de músicas de forró para surdos. Além disso, o desvio padrão da validação *k-fold* foi menor que 0,2% mostrando baixa variação de desempenhos entre as partições (*folds*).

O modelo de [25] foi avaliado somente para músicas sem ruído. A rede com melhor desempenho contou com 87 neurônios e obteve EPM=3,408%. Para o *dataset* sem ruído, o melhor modelo encontrado neste trabalho possui 48 neurônios na camada oculta e EPM=5,059%, que equivale a um aumento relativo no EPM de 48%. No presente estudo o *dataset* sem ruído foi acrescido de 42 músicas o que tornou o *dataset* atual mais genérico e mais complexo. A diversidade do novo *dataset* foi o principal motivo encontrado para justificar a aparente perda de desempenho do modelo, quando comparado com [25]. O *dataset* que o modelo avaliado apresentou menor erro foi o composto por áudios sem ruído e com ruído branco SB, com EPM=3,209%, enquanto que o *dataset* que o modelo apresentou pior desempenho foi o composto por músicas gravadas com ruído real R, com EPM=5,541%. A princípio, estes resultados dão a entender que apresentar arquivos com variados ruídos facilitaria o modelo identificar as componentes

principais, no entanto essa hipótese é inconsistente. O comportamento apresentado pelos modelos treinados com N e B, exibidos na Fig. 4, são praticamente coincidentes. O mesmo acontece com as curvas de erro dos modelos treinados com SR e RB.

Suspeita-se que o acréscimo de ruído branco aos arquivos sem ruído, não alterou de forma relevante as características espectrais na faixa de interesse e que os *datasets* S e B sejam muito parecidos. Portanto, uma amostra do *dataset* NB utilizada no treinamento pode ter uma amostra correspondente no teste, o que resultou no melhor desempenho exibido na Fig. 4.

B. Desempenho para o Cenário Real

Na Fig. 5, pode-se observar que o modelo cuja as amostras de treino pertenciam ao *dataset* R obteve o melhor desempenho quando avaliado com amostras diferentes do mesmo *dataset*. O EPM=5,541% foi significativamente menor que o dos demais ($p < 0.05$), mostrando que, para estimar a duração do compasso de músicas com ruído real, o melhor *dataset* deve conter estritamente músicas com ruídos reais.

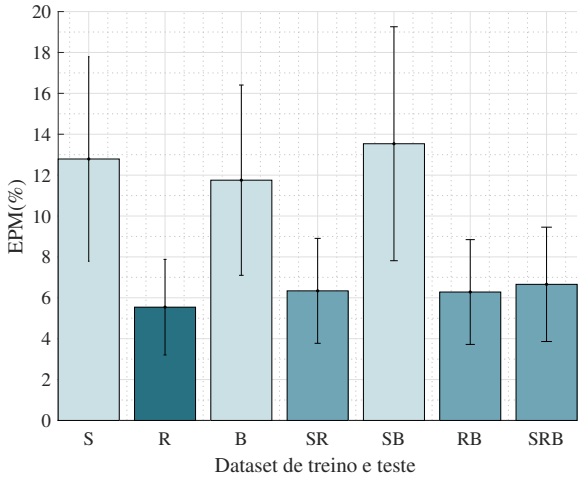
O modelo treinado com o *dataset* SR apresentou EPM=6,227% (Fig. 5a), desempenho que não é significativamente diferente dos modelos treinados com RB e SRB. Apesar de apresentar o segundo melhor desempenho, o modelo treinado com SR pode ser uma alternativa para aumentar o banco de dados, visto que as músicas sem ruído são mais fáceis de obter. Vale destacar que o ruído real, apresentado neste trabalho pelo *dataset* R, é específico de uma situação de dança que é muito ruidosa. É possível que em um cenário com menos ruído, o modelo treinado com SR possa ter maior capacidade de generalização.

Os *datasets* menos eficientes no treinamento dos modelos foram S, B e SB com EPM>11%, mostrando a necessidade de inserção de músicas com ruídos reais no banco de dados. Isto reforça a suspeita de que a inserção do ruído branco não fez mudanças significativas no *dataset* e que S e B são muito semelhantes.

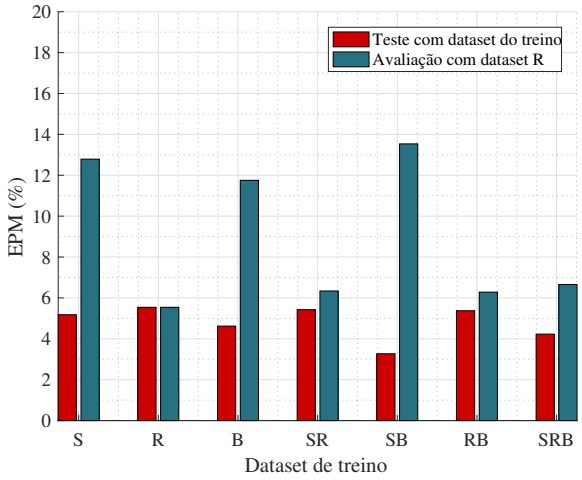
Na Fig. 5b, destaca-se a queda de desempenho de todos os modelos ao serem submetidos ao *dataset* exclusivo de ruídos reais. Esse fenômeno era esperado conforme discutido por [26], ressaltando a importância de utilizar bancos de dados reais para avaliar a aplicabilidade do modelo em mundo real.

O modelo que mais sofreu com os ruídos reais foi o treinado com SB que foi o modelo com menor EPM na etapa de teste. O erro deste modelo aumentou de 3,209% para 12,950%, que equivale à um aumento de aproximadamente 300%. Este aumento indica a baixa capacidade de generalização do modelo, fruto da pouca diversidade do *dataset* SB.

O modelo treinado com R não apresentou queda de desempenho (barra verde idêntica à vermelha na Fig. 5b) porque o *dataset* utilizado para ambos os testes foram os mesmos, conforme explicado anteriormente. Todos os modelos apresentados a amostras com ruído real durante o treino alcançaram erro médio absoluto menor que 100 ms, onde o modelo treinado com o dataset R, poderá passar o ritmo de músicas de forró para surdos com erro de 78 ms para cima ou para baixo do valor de compasso da música.



(a) Resultado dos modelos quando apresentados ao *dataset* R. As barras com cores diferentes possuem valores significativamente diferentes ($p < 0.05$).



(b) Comparação do desempenho obtido pelas melhores redes quando apresentadas ao *dataset* R em relação à etapa de teste.

Fig. 5. Desempenho dos modelos *dataset* e ruídos reais e sua comparação com os desempenhos obtidos na etapa de teste.

V. Desempenho do Estado da Arte em Estimação de Componentes Musicais

Na Tabela II são apresentados trabalhos que estimaram componentes musicais, as componentes estimadas por eles e os desempenhos dos melhores modelos. As componentes estimadas foram BPM e estrutura do compasso (EC). Para a Acurácia 1 e o F1, foram considerados acertos diferenças de até 4%.

O resultado do modelo proposto possui Acurácia1 inferior à todos os trabalhos revisados, os melhores desempenhos foram 79% para a estimação do BPM [27] e 74,5% para estrutura do compasso [29], conforme apresentado na Tabela II, sendo difícil inferir o desempenho final de uma eventual combinação dos modelos apresentados nesses trabalhos. Quando se observa as métricas de regressão, o modelo avaliado apresenta desempenhos mais otimistas, com EPM 5,541% e R2=0,771. No entanto, os trabalhos anteriores trataram a tarefa de estimação do

TABELA II
PARALELO DO DESEMPENHO DO MODELO CITADO COM A LITERATURA.

Ref	Saída	Acu1	F1	EPM	R2
[27]	BPM	**79**	-	-	-
[29]	EC	-	82	-	-
[30]	BPM	73,4	-	-	-
[31]	BPM	77,41	-	-	-
[33]	BPM	84,6	-	-	-
	EC	**74,5**	-	-	-
Avaliado	DC	48,035	64,897	5,541	0,771

BPM como uma tarefa de classificação binária, portanto, não foram apresentados desempenhos para métricas de regressão, impossibilitando uma comparação mais justa com a literatura.

VI. CONCLUSÃO

Este trabalho propôs a avaliação de um modelo baseado em rede PMC para futuramente embarcar um aplicativo que passe o ritmo de músicas de forró para surdos em um espaço de dança por meio de vibração. O modelo avaliado foi treinado com sete composições de *dataset* e a apresentou erro menor que 6% em todos os cenários. Os modelos que não foram apresentados à amostras com ruídos reais durante o treinamento foram significativamente mais afetados, com erro se aproximando a 15% quando testados com amostras reais. Os modelos que foram apresentados às músicas com ruídos reais durante o treino mantiveram erro inferior a 7% na fase de teste, mas foram inferiores ao modelo treinado somente com amostras com ruídos reais, que obteve erro significativamente menor de 5,541% que equivale a um erro médio absoluto de 78 ms.

Foi comprovada a capacidade do modelo avaliado de aprender a duração do compasso de músicas de forró em variados cenários, desde que tenha sido apresentado à amostras da mesma natureza durante o treino. Além disso, o ruído branco foi ineficaz para simular os ruídos de um espaço de dança. Portanto, o uso de músicas com ruídos reais no treinamento foi essencial para capacitar o modelo a estimar o compasso neste cenário, sendo o mais indicado para embarcar, futuramente, o aplicativo para auxiliar surdos a dançarem.

O presente trabalho apresenta duas fortes limitações. A primeira delas está relacionada ao número reduzido de músicas no banco de dados, principalmente do *dataset* com ruído real, que teve metade do tamanho. A fim de aumentar a diversidade das amostras de treino e melhorar a generalização do modelo treinado, é necessário acrescentar músicas no banco de dados. Além disso, uso de técnicas de *data augmentation* podem ser exploradas, bem como, acréscimo outras situações do cotidiano no *dataset* com ruídos reais, como músicas reproduzidas por computador ou celular em ambientes domésticos.

A segunda limitação consiste no método adotado para a seleção do número de neurônios e a validação do modelo. A abordagem proposta permitiu que amostras diferentes de uma mesma música fossem utilizadas nos *datasets* de treino, validação e teste. Desta forma, o modelo pode estar reconhecendo a música na fase de teste e estar associando com a duração do compasso das amostras que estiveram no treino. Neste caso, o modelo poderá ter desempenho menor quando apresentado a uma música que não teve amostras compondo os dados de treino. Para avaliar essa hipótese será necessário testar o modelo com músicas novas e avaliar possível queda de desempenho.

Em trabalhos futuros será necessário avaliar o quanto o erro obtido poderá atrapalhar a performance dos surdos dançando e qual limiar de erro será necessário alcançar para que a aplicação contribua de forma positiva para a experiência dos surdos com o forró.

AGRADECIMENTO

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. A Priscila da Silva Maradini e o projeto "Pé de Serra no Campus", pelas valiosas contribuições na criação do banco de dados.

REFERENCES

[1] WHO, "Deafness and hearing loss," mar 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss
[2] D. C. Baynton, *Forbidden signs: American culture and the campaign against sign language*. University of Chicago Press, 1996.
[3] M. A. L. Soares, *A educação do surdo no Brasil*. Editora Autores Associados, 2015.
[4] R. M. d. Quadros and L. B. Karnopp, *Língua de sinais brasileira: estudos lingüísticos*. Artmed, 2007.
[5] C. A. Padden and T. Humphries, *Inside deaf culture*. Harvard University Press, 2009.
[6] B. P. Tucker, "Deaf culture, cochlear implants, and elective disability," *Hastings Center Report*, vol. 28, no. 4, pp. 6–14, 1998.
[7] P. Kushalnagar, C. J. Moreland, A. Simons, and T. Holcomb, "Communication barrier in family linked to increased risks for food insecurity among deaf people who use american sign language," *Public Health Nutrition*, vol. 21, no. 5, pp. 912–916, 2018.
[8] J. Fellinger, D. Holzinger, and R. Pollard, "Mental health of deaf people," *The Lancet*, vol. 379, no. 9820, pp. 1037–1044, 2012.
[9] M. L. Fox, T. G. James, and S. L. Barnett, "Suicidal behaviors and help-seeking attitudes among deaf and hard-of-hearing college students," *Suicide and Life-Threatening Behavior*, vol. 50, no. 2, pp. 387–396, 2020.
[10] R. A. Cooper and R. Cooper, "Rehabilitation engineering: A perspective on the past 40-years and thoughts for the future," *Medical Engineering and Physics*, vol. 72, pp. 3–12, 2019.
[11] G. Raiola, "Inclusion in sport dance and self perception," *Sport Science*, vol. 8, no. 1, pp. 99–102, 2015.
[12] M. R. Zitomer, "Children's perceptions of disability in the context of elementary school dance education," *Revue phénEPS/PHEnex Journal*, vol. 8, no. 2, 2016.
[13] S. Koch, T. Kunz, S. Lykou, and R. Cruz, "Effects of dance movement therapy and dance on health-related psychological outcomes: A meta-analysis," *The Arts in Psychotherapy*, vol. 41, no. 1, pp. 46–64, 2014.
[14] S. C. Koch, R. F. F. Riege, K. Tisborn, J. Biondo, L. Martin, and A. Beelmann, "Effects of dance movement therapy and dance on health-related psychological outcomes. a meta-analysis update," *Frontiers in Psychology*, vol. 10, p. 1806, 2019.
[15] F. H. F. Wu and J. S. R. Jang, "A supervised learning method for tempo estimation of musical audio," in *2014 22nd Mediterranean Conference on Control and Automation*, 2014, pp. 599–604.
[16] E. Van Dyck, D. Moelants, M. Demey, A. Deweppe, P. Coussement, and M. Leman, "The impact of the bass drum on human dance movement," *Music Perception*, vol. 30, no. 4, pp. 349–359, 2013.
[17] M. Ezawa, "Rhythm perception equipment for skin vibratory stimulation," *IEEE Engineering in Medicine and Biology Magazine*, vol. 7, no. 3, pp. 30–34, 1988.
[18] Y. Dong, J. Liu, Y. Chen, and W. Y. Lee, "Salsaasst: Beat counting system empowered by mobile devices to assist salsa dancers," in *2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems*, 2017, pp. 81–89.

[19] H. Florian, A. Mocanu, C. Vlasin, J. Machado, V. Carvalho, F. Soares, A. Astilean, and C. Avram, "Deaf people feeling music rhythm by using a sensing and actuating device," *Sensors and Actuators A: Physical*, vol. 267, pp. 431–442, 2017.
[20] P. Tranchant, M. M. Shiell, M. Giordano, A. Nadeau, I. Peretz, and R. J. Zatorre, "Feeling the beat: Bouncing synchronization to vibrotactile music in hearing and early deaf people," *Frontiers in Neuroscience*, vol. 11, p. 507, 2017.
[21] A. Sharp, B. A. Bacon, and F. Champoux, "Enhanced tactile identification of musical emotion in the deaf," *Experimental Brain Research*, vol. 238, no. 5, pp. 1229–1236, 2020.
[22] E. Frid, "Accessible digital musical instruments-a review of musical interfaces in inclusive music practice," *Multimodal Technologies and Interaction*, vol. 3, no. 3, p. 57, 2019.
[23] A. C. d. Q. Junior, E. C. Fontes, R. Dias, and C. M. Volp, "Caracterização do xote e do baião dançados no interior do estado de são paulo," *Movimento*, vol. 15, no. 3, pp. 233–247, 2009.
[24] A. D. P. D. Santos, L. M. Tang, L. Loke, and R. Martinez-Maldonado, "You are off the beat! is accelerometer data enough for measuring dance rhythm?" in *ACM International Conference Proceeding Series*, 2018.
[25] L. F. Paiva, H. G. Lopes, L. B. Felix, and R. V. Neves, "Estimação do compasso musical do forró utilizando rede perceptron multicamadas," in *Anais do Congresso Brasileiro de Automática*, vol. 2, no. 1, 2020.
[26] C. W. Wu, C. Dittmar, C. Southall, R. Vogl, G. Widmer, J. Hockman, M. Muller, and A. Lerch, "A review of automatic drum transcription," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 26, no. 9, pp. 1457–1483, 2018.
[27] A. Eronen and A. Klapuri, "Music tempo estimation with k-nn regression," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 1, pp. 50–57, 2010.
[28] F. H. F. Wu and J. S. R. Jang, "A supervised learning method for tempo estimation of musical audio," in *22nd Mediterranean Conference on Control and Automation*, 2014, pp. 599–604.
[29] E. Quinton, M. Sandler, and C. Harte, "Extraction of metrical structure from music recordings," in *18th International Conference on Digital Audio Effects*, 2015, pp. 1–7.
[30] S. Böck, F. Krebs, and G. Widmer, "Accurate tempo estimation based on recurrent neural networks and resonating comb filters," in *16th International Society for Music Information Retrieval Conference*, 2015, pp. 625–631.
[31] H. Schreiber and M. Müller, "A post-processing procedure for improving music tempo estimates using supervised learning," in *18th International Society for Music Information Retrieval Conference*, 2017, pp. 235–242.
[32] S. Gulati, V. Rao, and P. Rao, "Meter detection from audio for indian music," in *Speech, Sound and Music Processing: Embracing Research in India*, vol. 7172, 2012, pp. 34–43.
[33] C. Uhle and J. Herre, "Estimation of tempo, micro time and time signature from percussive music," in *Proc. of the 6th Int. Conference on Digital Audio Effects*, 2003, pp. 1–6.
[34] A. C. d. Q. Junior and C. M. Volp, "Forró universitário: a tradução do forró nordestino no sudeste brasileiro," *Motriz*, vol. 11, no. 2, pp. 127–120, 2005.
[35] J. Roth, J. Ehlers, C. Getschmann, and F. Echtler, "Tempowatch: A wearable music control interface for dance instructors," in *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, 2021.
[36] I. N. d. Silva, D. H. Spatti, and R. A. Flauzino, *Redes Neurais Artificiais para Engenharia e Ciências Aplicadas*, 2nd ed. São Paulo: Artliber, 2016.

Lucas Ferreira Paiva recebeu o grau de Bacharel em Engenharia Elétrica pela Universidade Federal de Viçosa (UFV), Viçosa, Brasil, em 2020, é mestrando em Ciência da Computação na UFV, seus interesses de pesquisa são Inteligência Computacional e Processamento de Sinais.

Hugo Gonçalves Lopes é estudante de Engenharia Elétrica na Universidade Federal de Viçosa (UFV), seus interesses de pesquisa são Inteligência Computacional e Processamento de Sinais.

Elizabeth Regina Alfaro Espinoza possui graduação em Microbiologia y Parasitologia pela Universidad Nacional de Trujillo, Trujillo, Perú. Atualmente é doutoranda em Bioinformática pela Universidade Federal de Minas Gerais, Belo Horizonte, Brasil. Seus interesses de pesquisa incluem bioinformática estrutural e desenvolvimento de software para gerenciamento de dados biológicos.

Leonardo Bonato Félix possui graduação em Engenharia Elétrica pela Universidade Federal de São João del Rei (2002), mestrado e doutorado em Engenharia Elétrica pela Universidade Federal de Minas Gerais (2004 e 2006, respectivamente). Atualmente é professor associado da Universidade Federal de Viçosa, membro do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de São João del Rei e coordena o Núcleo Interdisciplinar de Análise de Sinais (NIAS/UFV). Atua nas área de Inteligência Computacional, Processamento de Sinais, Engenharia Biomédica, Instrumentação Eletrônica e Teoria da Detecção.

Rodolpho Vilela Alves Neves recebeu o grau de Bacharel em Engenharia Elétrica pela Universidade Federal de Viçosa (UFV), Viçosa, Brasil, em 2011, e os graus M.Sc. e D.Sc. em Engenharia Elétrica pela Universidade de São Paulo, São Carlos, Brasil, em 2013 e 2018, respectivamente. De 2015 a 2016, ele esteve como Pesquisador Visitante na Aalborg University, Dinamarca. Atualmente, é Professor Adjunto no Departamento de Engenharia Elétrica na UFV. Seus interesses de pesquisa incluem controle inteligente de sistemas dinâmicos e gerenciamento de microrredes de energia.