# Transcrição de Ritmos na Caixa da Bateria

**Thiago Martin Poppe** [1], **Flavio Vinicius Diniz de Figueiredo** [1]

[1]Instituto de Ciências Exatas – Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte - MG - Brasil

tmartinpoppe@gmail.com, flaviovdf@gmail.com

***Abstract.*** *The existing literature for automatic drum transcription is mainly focused on estimating the playing times of each piece of the drum kit. However, no studies were found that deal with the presentation of the results to an end user, typically a musician. In this sense, we propose the presentation of these results in the form of automatically generated scores from the estimation of the playing times emitted by a drum kit, more specifically for the snare drum. To do this, we use the pre-estimated playing times from the MDBDrums database along with meta information relevant to the song, such as the BPM and time signature. The product of this study proved to be promising, enabling new research fronts in this area.*

***Resumo.*** *A literatura existente para a transcrição automática de bateria é concentrada principalmente na estimação dos tempos de toque de cada peça do kit da bateria. Sendo assim, nenhum estudo foi encontrado que trata da apresentação dos resultados para um usuário final, tipicamente um músico. Neste sentido, propomos a apresentação destes resultados na forma de partituras geradas automaticamente a partir da estimação dos tempos de toque emitidos por um kit de bateria, mais especificamente para a caixa da bateria. Para tal, utilizamos os tempos de toque pré-estimados da base de dados MDBDrums juntamente com meta informações pertinentes à música, como o BPM e fórmula de compasso. O produto deste estudo provou-se promissor, viabilizando novas frentes de pesquisa voltadas para essa área.*

## 1. Introdução

Para amantes da música, uma grande dificuldade encontrada quando se quer tocar alguma canção com um instrumento específico é achar a partitura que diga como fazer isso. A partitura vai apresentar informações como o BPM (batidas por minuto) da música, quais notas tocar em qual tempo, articulações sonoras ao longo da música e, no caso da bateria por exemplo, até com qual mão tocar cada nota. Elaborar uma partitura apenas escutando uma música é um trabalho que demanda muito tempo e experiência do músico.

Uma partitura possui alguns componentes importantes, dentre eles podemos citar:

1. **Notas:** Situadas em linhas ou espaços ao longo da pauta, indicando qual nota devemos tocar e em qual tempo ela deve ser tocada;
2. **Clave:** Situada no começo da partitura, podendo também aparecer em outros trechos da partitura, indicando para o músico qual a referência que iremos utilizar para ler corretamente as notas na pauta;

3. **Fórmula de compasso:** Acompanha a clave, basicamente explicando como os compassos da partitura se comportam. Em outras palavras, de forma geral, expressará o "tamanho" dos compassos da partitura.

   Na Figura [1], podemos perceber alguns desses fragmentos citados:

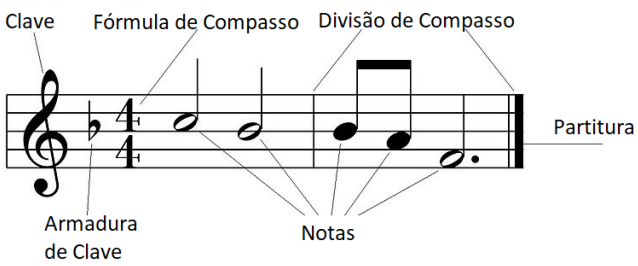


**Figura 1. Exemplo de uma partitura e seus principais componentes**

Nesse estudo, pretendíamos fazer uma primeira tentativa na tarefa de gerar partituras automaticamente dado o áudio de uma música. Por se tratar de tarefa muito complexa, que demanda a análise de muitos detalhes, como qual o BPM utilizado, qual nota foi tocada, qual a duração da nota e qual instrumento está sendo tocado, decidimos quebrar esse problema em partes menores e mais simples, possibilitando que esforços futuros gradativamente avancem em direção à resolução do problema completo.

Dessa forma, o objetivo do nosso trabalho foi a criação de uma partitura para ritmos gerados apenas por uma caixa de bateria. Aqui, nos preocupamos apenas em definir corretamente o BPM da música, estimar a sua fórmula de compasso e definir as figuras de tempo das notas dependendo do ritmo em que são tocadas as batidas. A estimação dos tempos de toque da caixa da bateria servirão de entrada para o algoritmo proposto e serão, por ora, recuperados através dos dados pré-anotados da base de dados MDBDrums [Southall et al. 2017].

O resto desse trabalho será estruturado da seguinte forma: a seção 2 oferece uma análise de trabalhos relacionados; a seção 3 apresenta a metodologia utilizada, incluindo detalhes da base de dados utilizada e a solução proposta. Por fim, experimentos e resultados podem ser encontrados na seção 4, seguidos por uma conclusão na seção 5.

## 2. Trabalhos Relacionados

A literatura da área de processamento de aúdios de bateria é bastante extensa, porém, como mencionado anteriormente, aparentemente não há muitos estudos que focam especificamente na parte da transcrição de partituras.

Um trabalho muito interessante publicado na conferência ISMIR, explora a utilização de uma rede neural recorrente bidirecional para extrair as informações de batidas da caixa, bumbo e chimbal (hihat) do kit de bateria [Southall et al. 2016]. Neste trabalho os autores conseguiriam, de forma satisfatória, segmentar os momentos da música associados com cada um dos instrumentos presentes no kit da bateria. Assim, mesmo que uma partitura não tenha sido gerada como produto final deste estudo, é possível encontrar trabalhos em direção à uma solução satisfatória para o problema da transcrição da bate-

ria. Tais trabalhos serão futuramente analisados e possivelmente incorporados durante o desenvolvimento da pesquisa no POC II.

Outro trabalho bem interessante, tenta resolver o problema de transcrição utilizando modelos baseados em redes neurais convolucionais recorrentes [Vogl et al. 2017]. A principal contribuição do artigo é a inclusão de meta informações normalmente desconsideradas em outros trabalhos, como a duração de um compasso, o tempo (ou BPM) e a fórmula de compasso. Com ela, o método identifica e faz uso de informações pertinentes à rítmica da música, fortemente correlacionadas com a transcrição da bateria.

## 3. Metodologia

Ao longo da literatura de Recuperação de Informação Musical, o problema de transcrição foi definido como o processo de estimar os tempos de toque de cada peça da bateria. Para esse estudo, iremos definir o problema de transcrição tal qual é definido na área musical, ou seja, dado uma música (como por exemplo um áudio em formato `.wav`) iremos gerar uma partitura dos instrumentos que compõem aquela música. No nosso caso, dado um áudio de uma bateria, iremos escrever a partitura exclusivamente da caixa da bateria.

Na seção 3.1, iremos explicar um pouco mais em detalhes a base de dados utilizada para o projeto. Já nas seções seguintes, iremos abordar os componentes necessários para o entendimento do algoritmo proposto, explicado em mais detalhes na seção 3.3.

### 3.1. Base de dados utilizada

Durante o desenvolvimento da pesquisa, bem como para a etapa dos experimentos, foi utilizado a base de dados **MDBDrums** [Southall et al. 2017], que representa um subconjunto de áudios da base **MedleyDB** [Bittner et al. 2014], composta por áudios de bateria, bem como anotações dos tempos em que cada peça da bateria é tocada. Como o foco da pesquisa é trabalhar diretamente com a caixa da bateria, foi feito uma filtragem das anotações referentes apenas a ela.

A base **MDBDrums** possui ao todo 23 tracks com um total de 7994 eventos de *onset*, ou seja, de inícios de uma nota musical ou de um som. Além disso, a base de dados cobre uma alta gama de estilos musicais, como por exemplo: Rock, Country, Disco, Reggae e Jazz, providenciando assim um espaço amostral mais fidedigno do mundo real, permitindo com que diversas análises e tarefas sejam feitas sobre ela.

### 3.2. Extração das características sonoro/musicais

Inspirados pela proposta do estudo [Vogl et al. 2017], onde foi utilizado *features* pertinentes da música para construir a solução proposta, iremos extrair e agregar ao algoritmo final características sonoro/musicais dos áudios presentes na base de dados. Como características pertinentes para esse trabalho, temos: (i) o BPM, ou batidas por minuto da música, também chamado de *tempo*, que indica para nós o andamento da música; e (ii) a fórmula de compasso, que explicita como os compassos da partitura se comportam, i.e. quantas notas teremos em cada compasso da música.

#### 3.2.1. Estimação do *tempo*

Para estimar o BPM da música, utilizamos a biblioteca `librosa`, que oferece diversas funcionalidades para manipulação e extração de *features* de áudios. Em particular, foi uti-

lizado a função `beat.tempo`, que permite a extração do *tempo* da música através de uma abordagem via programação dinâmica e técnicas já consolidadas na área de Recuperação de Informação Musical.

Para proporcionar resultados mais interessantes, foi feito um ajuste de parâmetros da função utilizada. Em particular, foi utilizado uma agregação da estimação dos *tempos* de cada compasso através de uma média, bem como um *hop size*, ou seja, o tamanho do "passo" que daremos na análise do espectrograma do sinal, igual à 256 *frames*.

### 3.2.2. Detecção da fórmula de compasso

A estimação da fórmula de compasso de uma música é uma tarefa árdua, ainda mais se tratando de músicas com diversas mudanças rítmicas ou que utilizam fortemente de polirritmia, definida pelo uso simultâneo de duas ou mais estruturas rítmicas diferentes na mesma música, muito utilizada em gêneros progressivos, mas que também pode ser notado nos minutos iniciais da *Terceira Sinfonia* de Beethoven.

A fórmula de compasso é representada através de uma fração, como, por exemplo: $\frac{4}{4}$, $\frac{7}{8}$ e $\frac{5}{4}$, na qual o denominador representa uma nota de referência, seguindo a escala que pode ser observada na Figura [2]; e o numerador quantas notas de referência teremos por compasso. Vale ressaltar que podemos ter subdivisões das notas de referência, ou seja, em um compasso $\frac{4}{4}$ podemos ter 4 semínimas ou 8 colcheias, uma vez que 1 colcheia representa a metade da duração de uma semínima.

| Nome | Formato da Nota | Pausa Equivalente | Duração Rítmica | Nº da Nota de Referência |
|---|---|---|---|---|
| Semínima |  |  | 1 batida | 4 |
| Colcheia |  |  | 1/2 batida | 8 |
| Semicolcheia |  |  | 1/4 batida | 16 |
| Fusa |  |  | 1/8 batida | 32 |

**Figura 2. Escala das notas de referência**

Para realizar a detecção da fórmula de compasso das músicas, optamos por implementar uma solução já publicada que se baseia em uma análise da matriz de similaridade da música [Coyle and Gainza 2007]. Matrizes de similaridade são altamente utilizadas no cenário de Recuperação de Informação Musical, principalmente por assumirmos que músicas no geral possuem tendências que se repetem ao longo do tempo [Foote et al. 2001]. Tal solução é ao mesmo tempo simples e robusta, garantindo resultados satisfatórios para a base de dados utilizada. A única modificação feita na publicação original, é que para esse trabalho, não foi implementanda a detecção de anacruse, visto que as músicas utilizadas para essa pesquisa não possuem essa característica musical.

## 3.3. Algoritmo de transcrição proposto

Em linhas gerais, o algoritmo final de transcrição proposto possui as seguintes etapas: (i) recuperação das características sonoro/musicais do áudio, i.e. BPM e fórmula de com-

passo através dos métodos citados anteriormente; (ii) computar a duração em segundos da nota de referência, para auxílio de referência rítmica durante o processo de transcrição; e, finalmente, (iii) através dos tempos de toque estimados *a priori*, executamos um algoritmo para estimar as figuras de tempo de cada toque da caixa da bateria.

### 3.3.1. Duração em segundos da nota de referência

Para computar a duração em segundos da nota de referência, utilizamos a seguinte fórmula matemática, sendo $tempo$ o BPM estimado da música e "reference note" o denominador da fórmula de compasso, que sempre será um número múltiplo de 2.

$$\text{step} = \frac{60}{tempo} * \frac{4}{\text{reference note}}$$

A intuição por trás da fórmula consiste em: subdividirmos 1 minuto (60 segundos) de acordo com o BPM estimado, recuperando assim a duração em segundos de uma batida. Posteriormente, ajustamos esse valor de acordo com a subdivisão utilizada como nota de referência, uma vez que o BPM da música é expresso levando em consideração a figura de tempo de uma semínima. Em outras palavras, se estivermos trabalhando com uma nota de referência igual a uma colcheia, devemos dividir a duração de uma batida pela metade, já que a duração rítmica de uma colcheia equivale à metade da duração rítmica de uma semínima, como podemos observar novamente na Figura [2], da seção 3.2.2.

### 3.3.2. Estimação das figuras de tempo

Tendo em vista todos os componentes explicados anteriormente, juntamente com os tempos de toque da caixa, estimados *a priori*, o algoritmo para a estimação das figuras de tempo seguirá os seguintes passos:

1. Inicialize um acumulador com o valor 0 e um ponteiro para a posição inicial de uma lista contendo os tempos de toque da caixa;
2. Compute a diferença entre o acumulador e o tempo da caixa atual, limitando ela ao valor de $step$, computado da mesma forma demonstrada na seção 3.3.1. Essa diferença irá representar a duração em segundos da nota tocada;
3. Subdivida o valor de $step$ de acordo com as figuras de tempo a serem analisadas e encontre aquela que mais se aproxima da duração da nota estimada no passo 2. Por exemplo, se a nota de referência for uma semínima, e o valor de $step$ for igual à 1 segundo, a figura de tempo de uma semínima terá uma duração de 1 segundo, uma colcheia terá uma duração de $\frac{1}{2}$ segundo, e assim por diante;
4. Caso o valor do acumulador seja menor que o tempo da caixa atual, iremos escrever uma pausa na partitura de mesma figura de tempo estimada no passo anterior. Caso contrário, iremos escrever uma nota. Em ambos os casos, iremos atualizar o acumulador com a duração em segundos da nota/pausa estimada e, somente se escrevermos uma nota, iremos avançar o ponteiro para a próxima posição;

5. Se todos os tempos de caixa foram processados, vá para o passo 6. Senão, volte para o passo 2;
6. A duração da última nota não poderá ser estimada da mesma forma que no passo 2. Para isso, estime a sua duração como a diferença entre o último tempo de caixa e o tempo final do compasso que ela está inserida. Com essa estimação em mãos, repita os passos 3 e 4 até que o acumulador seja maior ou igual que o tempo final do compasso atual.

Por mais que o algoritmo enunciado anteriormente seja relativamente simples e compacto, ele possui algumas nuances. A primeira que podemos observar é que no passo 2, iremos sempre limitar a duração da nota pelo valor de $step$. Essa é uma simplificação que podemos assumir ao trabalharmos com instrumentos de altura indefinida, como o caso da bateria. Em instrumentos de altura indefinida, ao contrário de instrumentos de altura definida, nós não temos uma noção de sustentação do som. Em outras palavras, assumindo que uma nota não interrompa prematuramente a duração do som de outra, a duração sonora de qualquer figura de tempo será a mesma. Nós podemos observar esse fenômeno na Figura [3], onde na parte superior nós temos um espectrograma de uma bateria tocando 3 notas, e na parte inferior um piano (instrumento de alturas definidas) sustentando uma única nota na mesma duração dos 3 toques da bateria.

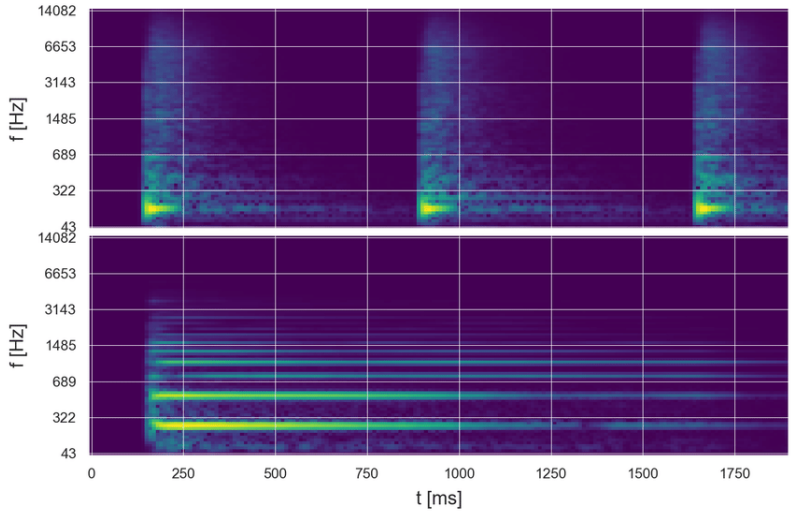


**Figura 3. Espectrograma comparando a duração de notas na bateria e no piano**

Um outro ponto que vale mencionar no algoritmo de estimação das figuras de tempo, é o cálculo do tempo final do compasso em que a última nota está inserida. Esse valor pode ser facilmente obtido uma vez que sabemos o valor de $step$, bem como a fórmula de compasso. Como o valor de $step$ representa a duração da nota de referência, sabemos que um compasso possui uma duração de $n * step$ segundos, onde $n$ representa o numerador da fórmula de compasso, ou seja, quantas notas de referência teremos por compasso. Com essa informação em mãos, basta somarmos esse valor $X$ vezes até que ele ultrapasse o último tempo de caixa estimado, indicando assim o fim do último compasso da música.

Por fim, como a estimação dos tempos de caixa possui algum erro associado, iremos incluir uma margem de erro $\epsilon$ pequena nas desigualdades feitas durante o passo

[4]. Particularmente, um bom valor de $\epsilon$ observado na prática estaria um pouco abaixo da duração da menor figura de tempo levada em consideração durante o passo [3]. Por exemplo, se a menor figura de tempo analisada for uma fusa, e estivermos trabalhando com uma nota de referência igual à uma semínima e $step$ igual à 1 segundo, teremos que $\epsilon < 1/8$ segundos.

## 4. Análises e Experimentos

O método descrito na seção 3.3 foi implementado na linguagem de programação Python 3.9.12. Os experimentos foram executados tanto localmente, em uma máquina Windows através do WSL (*Windows Subsystem for Linux*) com um processador Intel Core i7 7ª geração, 16GB de memória RAM e 1TB de HD; quanto através do Google Colab.

Além disso, as partituras foram escritas utilizando a biblioteca `music21` [Cuthbert and Ariza 2010], desenvolvida por pesquisadores do MIT e disponibilizada de forma gratuita.

### 4.1. Exemplo de transcrição

Para exemplificar o resultado obtido pelo algoritmo proposto na seção 3.3, iremos utilizar a música *Welcome To The Black Parade*, da banda *My Chemical Romance*, que não está presente na base de dados utilizada, porém apresenta um padrão rítmico no início da música bem interessante, baseado fortemente em ritmos tocados em marchas. Para esse exemplo, os tempos de caixa foram estimados manualmente e podemos observar a partitura original do trecho em questão na Figura [4].

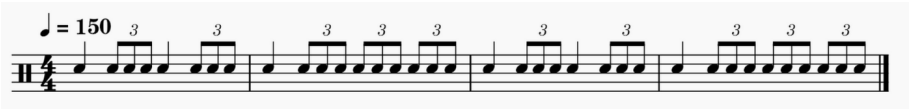


**Figura 4. Transcrição original da bateria de *Welcome To The Black Parade***

Podemos observar que o trecho extraído possui figuras rítmicas relativamente complexas, como a presença de quiálteras (mais especificamente tercinas), que são subdivisões de uma nota em mais ou menos partes do que o normal (múltiplos de 2). Ao realizarmos uma transcrição automatizada de acordo com o algoritmo proposto, temos o resultado demonstrado através da Figura [5], onde temos uma transcrição completamente errada do ritmo da música original. Porém, podemos perceber que o motivo desse erro está associado com a estimação errônea do BPM da música, onde estimamos um BPM igual à 112 batidas por minuto; enquanto que a música original possui 150 batidas por minuto. Isso pode ser explicado por não termos muita variação rítmica e por ser um trecho extremamente pequeno, cerca de 8 segundos. Além de não termos o contexto da música como um todo ao fazer a análise, o que torna o método de estimação do BPM impreciso.

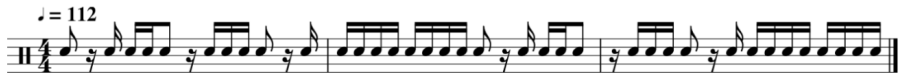


**Figura 5. Transcrição automatizada da bateria de *Welcome To The Black Parade***

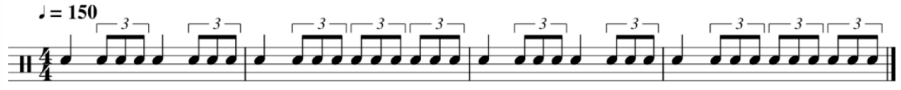


**Figura 6. Transcrição automatizada utilizando BPM correto**

Podemos perceber na Figura [6] que, ao corrigirmos o BPM estimado, temos uma transcrição perfeita, indicando assim que o método proposto também consegue lidar com figuras de tempo mais complexas.

Além disso, um resultado interessante é que temos uma certa tolerância ao erro associado com a estimação do BPM, como mostrado na Figura [7]. Podemos observar que um erro de até 15 batidas por minuto abaixo do valor real ainda resulta em uma transcrição da música de forma perfeita. Porém, a margem de erro acima do valor real é bem menor, já que teremos uma estimação da duração da nota maior do que o que ela realmente é, resultando em algumas desigualdades darem falsos positivos e negativos.

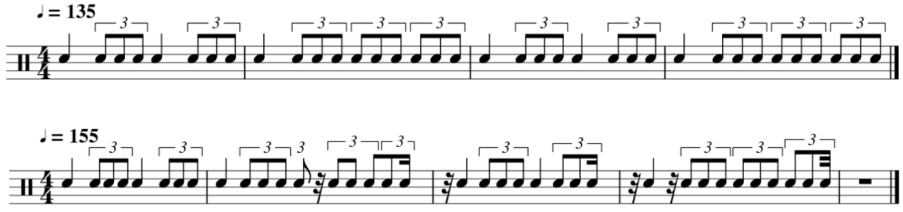


**Figura 7. Transcrição automatizada utilizando BPM = 135 e BPM = 155**

### 4.2. Resultados principais

Os resultados obtidos sobre a base de dados MDBDrums foram restringidos para os 10 estilos musicais mais diversos da base por motivos de visualização dos resultados. Porém, de forma geral, os valores observados nessa seção são similares para as demais instâncias da base. Além disso, foi utilizado ao longo do algoritmo de transcrição, as figuras de tempo mais utilizadas em partituras voltadas para bateria, ou seja: semínimas, colcheias, semicolcheias, fusas e tercinas.

Podemos observar pela Tabela [1], que o erro médio dos tempos das caixas transcritos está bem baixo, como retratado pela última coluna da tabela. Esses erros foram computados através da diferença média entre os tempos da caixa transcritos (os tempos de cada nota na partitura construída) com os tempos pré-anotados disponibilizados pela base. Em particular, podemos perceber que a última linha da tabela, relacionada com o estilo musical Zeppelin (referenciando a banda de rock dos anos 70, Led Zeppelin), possui o menor erro médio observado em toda a tabela, mesmo com uma estimação errada da fórmula de compasso. Uma possível explicação para isso seria de que a fórmula de compasso estimada (11/8) está bem próxima da fórmula de compasso 12/8, que pode ser vista como uma fórmula equivalente, ou uma "outra visão", da fórmula de compasso 4/4, que seria a fórmula correta da música.

Um outro ponto interessante de se notar é de que todas as músicas da base possuem uma fórmula de compasso relativamente simples, i.e. 4/4. Porém, ao escutarmos os estilos musicais presentes, podemos perceber que mesmo assim as músicas que compõem a base

possuem níveis diversificados de complexidade rítmica, no caso a utilização de figuras de tempo variadas.

| Song Style | Estimated Tempo | Correct Tempo | Estimated Time Signature | Correct Time Signature | Snare Times Mean Error |
|---|---|---|---|---|---|
| 80s Rock | 110 | 110 | 4/4 | 4/4 | 0.02002 |
| Beatles | 110 | 110 | 4/4 | 4/4 | 0.02042 |
| Disco | 110 | 110 | 4/4 | 4/4 | 0.01808 |
| Funk Jazz | 110 | 110 | 4/4 | 4/4 | 0.01159 |
| Grunge | 110 | 110 | 4/4 | 4/4 | 0.06829 |
| Hendrix | 110 | 110 | 4/4 | 4/4 | 0.01844 |
| Punk | 148 | 150 | 4/4 | 4/4 | 0.03569 |
| Rockabilly | 110 | 110 | 4/4 | 4/4 | 0.02818 |
| Speed Metal | 110 | 110 | 4/4 | 4/4 | 0.04913 |
| Zeppelin | 110 | 110 | 11/8 | 4/4 | 0.00971 |

**Tabela 1. Tabela de resultados para a base MDB Drums**

Porém, mesmo com um erro baixo como apontado pela tabela, ao observarmos a transcrição gerada para o estilo musical Beatles, percebemos que a partitura foi escrita de forma inusitada, possuindo pausas em quiálteras com valores fora do comum, destacadas pelas setas azuis na Figura [8]. O motivo dessas quiálteras estarem presentes seria por conta das duas notas na caixa, destacadas pelos círculos vermelhos, que acontecem ao longo da música.

Na gravação, o baterista faz um rudimento chamado *flam*, que é quando o baterista toca duas notas na caixa da bateria com as duas mãos quase ao mesmo tempo, onde uma delas tem o som mais fraco que a outra. Como não analisamos rudimentos no algoritmo de transcrição, ele irá estimar essa região como sendo dois toques separados na caixa da bateria, resultando em atrasos que tentam ser corrigidos pelas pausas.

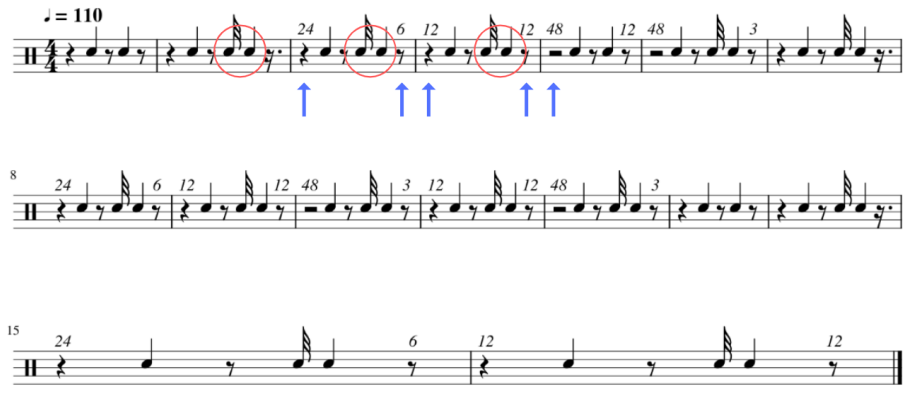


**Figura 8. Transcrição automatizada do estilo Beatles**

## 5. Conclusão

Nesse trabalho, utilizamos técnicas para a transcrição de partituras a partir de músicas que representam ritmos tocados no kit da bateria, em especial na caixa da bateria. Os

resultados obtidos demonstraram que as técnicas aplicadas foram satisfatórias, sobretudo devido à qualidade das estimações *a priori* dos tempos de toque anotados presente na base MDBDrums.

Conclui-se que o objetivo do trabalho foi alcançado, abrindo portas para trabalhos futuros, como por exemplo, a geração de partituras para músicas contendo mais de uma peça da bateria, como por exemplo: caixa, bumbo e chimbal, que será explorado futuramente ao longo da pesquisa a ser desenvolvida no POC II, juntamente com um método para a estimação dos tempos de toque de cada peça do kit de bateria. Além disso, podemos estender o trabalho feito considerando também alguns rudimentos, como é o caso do *flam*, durante a estimação das figuras de tempo de cada toque na caixa da bateria.

## Referências


Bittner, R., Salamon, J., Tierney, M., Mauch, M., Cannam, C., and Bello, J. (2014). Medleydb: A multitrack dataset for annotation-intensive mir research.

Coyle, E. and Gainza, M. (2007). Time signature detection by using a multi-resolution audio similarity matrix. *Journal of the Audio Engineering Society*.

Cuthbert, M. and Ariza, C. (2010). Music21: A toolkit for computer-aided musicology and symbolic music data. pages 637–642.

Foote, Jonathan, Cooper, M., and Matthew (2001). Visualizing musical structure and rhythm via self-similarity.

Southall, C., Stables, R., and Hockman, J. (2016). Automatic drum transcription using bi-directional recurrent neural networks. In *ISMIR*.

Southall, C., Wu, C.-W., Lerch, A., and Hockman, J. (2017). MDB drums: An annotated subset of medleydb for automatic drum transcription. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*.

Vogl, R., Dorfer, M., Widmer, G., and Knees, P. (2017). Drum transcription via joint beat and drum modeling using convolutional recurrent neural networks. In *ISMIR*.