CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

# Análise e classificação das segmentações de músicas

Nathália de Vasconcelos Silva

João Pessoa, PB

Junho – 2023

Nathália de Vasconcelos Silva

# Análise e classificação das segmentações de músicas

Monografia apresentada ao curso Engenharia de Computação do Centro de Informática, da Universidade Federal da Paraíba, como requisito para a obtenção do grau de Bacharela em Engenharia de Computação.

Orientadora: Prof.ª Dr.ª Thaís Gaudencio do Rêgo.
Coorientador: Prof.ª Dr. Yuri de Almeida Malheiros Barbosa.

João Pessoa, PB
Junho - 2023

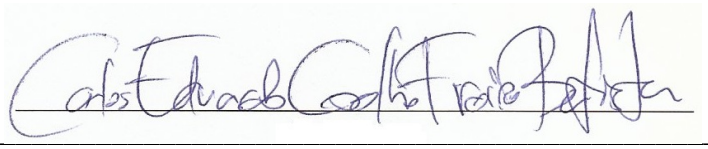
(Folha para o catálogo da Biblioteca)

CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

Trabalho de Conclusão de Curso de Engenharia de Computação intitulado **Análise e classificação das segmentações de músicas** de autoria de Nathália de Vasconcelos Silva, aprovada pela banca examinadora constituída pelos seguintes professores:

Prof.ª Dr.ª Thaís Gaudencio do Rêgo
Universidade Federal da Paraíba - UFPB

Prof. Dr. Yuri de Almeida Malheiros Barbosa
Universidade Federal da Paraíba - UFPB



Prof. Dr. Carlos Eduardo Coelho Freire Batista
Universidade Federal da Paraíba - UFPB

João Pessoa, 20 de Junho de 2023

# AGRADECIMENTOS

Agradeço especialmente aos meus pais, Lúcia Vasconcelos e Osmar Silva, por terem me proporcionado a possibilidade de alcançar os meus sonhos e os meus objetivos através da educação. Agradeço ao meu noivo, que porventura é colega de profissão e também de graduação, Janse Brasileiro, pelo apoio incondicional no dia a dia e por todos os momentos investidos nos estudos para a conclusão das nossas graduações em Computação. Agradeço à minha irmã Thaís Vasconcelos e aos meus amigos mais próximos, pelas palavras de incentivo para concluir a minha graduação e entregar este trabalho.

Gostaria de dedicar os meus sinceros agradecimentos aos professores que não só contribuíram intelectualmente para a minha jornada até então, mas também para os que foram exemplo extraclasse, e, antes de tudo, humanos e compreensivos com os seus alunos e suas eventuais limitações. Neste quesito, a professora Thaís Gaudencio se destaca. Este trabalho de conclusão de curso foi viabilizado por sua orientação. À ela, minha eterna gratidão. Também não posso deixar de citar os professores Protásio de Souza e Fernando Matos, que me forneceram oportunidades durante a graduação e despertaram o meu interesse pela pesquisa científica, tanto na Engenharia Elétrica, quanto em Computação.

# RESUMO

A análise estrutural de músicas possui dois subproblemas: detecção de limites temporais e o agrupamento estrutural, analisados aqui, detalhadamente. O resultado do primeiro subproblema influencia diretamente no resultado do segundo. Para automatizar estas atividades, foi utilizado, neste trabalho, o *framework* MSAF (do inglês, *Music Structure Analysis Framework*), com dois algoritmos, *Convex Non-Negative Matrix Factorization* (do português, Fatoração de Matriz Convexa Não-Negativa) e 2D *Fourier Magnitude Coefficients* (do português, Coeficientes de Magnitude de Fourier 2D), que utilizam a técnica de agrupamento *k-means*. Neste trabalho, além da performance do MSAF ter sido analisada, foi avaliado o aplicativo de música Moises. Foi construído, para tanto, um conjunto de dados de 5 músicas, padrão-ouro, rico em detalhes nas anotações das segmentações de músicas, que contaram com a avaliação de um músico profissional e da autora, estudante de teoria e prática musical com foco em guitarra elétrica. Para analisar os resultados obtidos, foram avaliadas principalmente as métricas PWF (do inglês, *Pairwise Frame Clustering*), além da medida-F. Foi feita uma análise comparativa entre os resultados dos limites temporais e dos rótulos sugeridos (agrupamento estrutural) das segmentações resultantes do *framework* MSAF e do aplicativo de música Moises, que divergem do padrão-ouro. A média do resultado do PWF nas 5 músicas do conjunto de dados para o Moises foi de 42,29%, com o MSAF, algoritmo 2DFMC, 50,08%. Por fim, para o algoritmo CNMF do MSAF, o PWF alcançou 44,18%. Concluiu-se que as divergências de notações utilizadas nas anotações influenciam diretamente no resultado final alcançado pelas aplicações, o que reforça a complexidade da atividade para lidar com a subjetividade ao realizá-la, além de evidenciar que há discordância entre os músicos avaliadores dos conjuntos de dados.

**Palavras-chave**: Análise Estrutural de Músicas. Técnica de Agrupamento. Segmentação de Músicas. Processamento de Sinais de Áudio.

# ABSTRACT

The structural analysis of music contains two subproblems: boundary detection and structural grouping, which are analyzed in this work in detail. The result of the first subproblem directly affects the second one. To automate these activities, the Music Structure Analysis Framework (MSAF) was used with two algorithms, Convex Non-Negative Matrix Factorization and 2D Fourier Magnitude Coefficients, which execute these two activities with the k-means clustering technique. In this work, beyond the analysis of MSAF's performance, the music application Moises was used. In this work, a small, but rich in details, dataset was built, which relied on the assessment of a professional musician and the author, a student of theoretical and practical music specializing in electric guitar. To analyze the obtained results, the main metrics were Pairwise Frame Clustering (PWF) and F-measure. A comparative analysis was made between the results of the boundary detections and the suggested labels (structural grouping) of the resulting segmentations of the MSAF framework and the Moises app, which diverge from the golden pattern. The mean of the result of PWF in 5 songs of the dataset for Moises was 42,29%, with MSAF, algorithm 2DFMC, 50,08%. Lastly, for the CNMF algorithm in MSAF, PWF reached 44,18%. It was concluded that the divergences of the notations adopted influenced directly in the final result achieved by the applications, regardless of being the framework or the Moises app, which reinforces the complexity of the activity to handle its subjectivity, besides evidencing the disagreement between the evaluating musicians of the datasets.

**Keywords:** Music Structure Analysis. Clustering Technique. Music Segmentation. Audio Signal Processing.

# LISTA DE ABREVIATURAS

2DFMC - *2D Fourier Magnitude Coefficients* (do português, Coeficientes de Magnitude de Fourier 2D)

CNMF - *Convex Non-Negative Matrix Factorization* (do português, Fatoração de Matriz Não-negativa Convexa)

CSV - *Comma-separated values* (do português, valores separados por vírgulas)

FFT - *Fast Fourier transform* (do português, transformada rápida de Fourier)

IDE - *Integrated Development Environment* (do português, ambiente integrado de desenvolvimento)

MIR - *Music Information Retrieval* (do português, pesquisa de informática musical)

MP3 - MPEG-1 Audio Layer 3

MSAF - *Music Structure Analysis Framework* (do português, *framework* de análise de estrutura musical)

PCP - *Pitch Class Profiles* (do português, conjunto de afinações)

PWF - *Pairwise Frame Clustering* (do português, Agrupamento de Pares Similares de Quadros)

SSM - Matriz de auto similaridade (do inglês, *Self-similarity Matrix*)

# LISTA DE FIGURAS

# LISTA DE TABELAS

# SUMÁRIO

# 1 INTRODUÇÃO

Nos últimos tempos, o campo da pesquisa de informática musical (do inglês, *Music Informatics Research* - MIR) recebeu contribuições relevantes no que se refere às técnicas utilizadas para realizar as atividades de segmentação musical, com o uso de técnicas de aprendizagem de máquina, maior poder computacional e conjuntos de dados maiores e mais ricos em informações. Uma das tarefas relacionadas ao MIR e que sofreu um forte interesse na automatização de tarefas foi a segmentação de músicas, pois estas informações podem produzir mapeamentos de alto nível das músicas, e podem ser aplicadas para melhorar a experiência de usuário em catálogos musicais na hora de entender, navegar e explorar grandes coleções de músicas (Nieto et. al., 2020).

As descobertas desta área podem resultar em produtos que estimulem a criatividade e a produção musical. Elas podem auxiliar músicos profissionais e amadores no que diz respeito ao fornecimento da análise mais específica dos segmentos de músicas. O mercado da música também pode se beneficiar com os resultados trazidos por esta pesquisa, pois os usuários finais podem receber recomendações baseadas nas estruturas musicais que sejam de sua preferência. A produção musical tem muito a ganhar também, pois ter uma biblioteca de músicas e ideias de músicas separadas por estruturas similares, pode auxiliar no processo criativo de compositores, além de ajudá-los a perceber padrões que possam vir a acontecer em músicas que tenham um alto grau de popularidade.

O tema também possui relevância educacional, pois os musicólogos podem utilizar o recurso automatizado de segmentação de músicas para estudar de forma estruturada as partes das músicas. Ele também pode ser utilizado no processo de colaboração com outros músicos, a fim de facilitar a comunicação e o entendimento sobre as segmentações das músicas e os seus detalhes acústicos.

## 1.1 PROBLEMÁTICA

O processo de análise e segmentação estrutural de músicas é custoso e demorado, além de necessitar de uma avaliação de um músico especialista. De acordo com Davies et. al. (2009), são necessários dados musicais para executar experimentos; verdades fundamentais

para comparar com os resultados dos algoritmos que estão sendo testados; uma forma significativa de mensurar a performance entre a saída do algoritmo e a verdade fundamental. Para as atividades relacionadas à MIR, os problemas de avaliação podem surgir em todas essas etapas.

Sobre os dados musicais, a distribuição dos sinais de áudios pode infringir direitos autorais dos artistas, logo, é um desafio ter dados *open source* em relação às músicas, o que acarreta em pesquisadores utilizarem conjuntos de dados de músicas populares, que são fáceis de encontrar no mercado da música. A desvantagem disso é realizar análises e desenvolvimento de algoritmos que apresentam boa performance em estilos musicais populares, mas que nem sempre apresentam bons resultados nos demais estilos musicais.

Já sobre as verdades fundamentais, elas demandam tempo e disponibilidade de músicos especialistas para realizarem a análise e anotação das segmentações. Sobre este ponto, também há um grande desafio na área, pois é uma atividade subjetiva, que conta com a percepção do músico e também a notação utilizada para anotar as segmentações das músicas varia de músico para músico. Normalmente, um músico com formação clássica anota as músicas em um formato diferente de um músico popular, e assim vão surgindo as divergências para a rotulação de segmentações de músicas.

Por fim, sobre a forma significativa de mensurar a performance dos algoritmos, o desafio é conseguir lidar adequadamente com a subjetividade e a ambiguidade, enquanto é provida uma forma de mensurar a performance do algoritmo que seja significativa e fácil de interpretar (Davies et. al., 2009).

### 1.1.1 Justificativa

A área de recuperação de informação de música necessita de conjuntos de dados confiáveis e curados por músicos especialistas. Devido aos direitos autorais para distribuição de sinais de áudio, muitos conjuntos de dados são disponibilizados sem a referência de áudio, apenas com as verdades fundamentais dos anotadores, o que dificulta o processo de verificação dessas anotações. Para realizar pesquisa e comparar técnicas de segmentação estrutural de músicas, é importante ter conjuntos de dados que foram produzidos por pessoas com conhecimento teórico musical. Tendo em vista que as pesquisas são feitas com conjuntos de dados populares na área, este trabalho visa disponibilizar uma nova base de dados[1]

[1] Base de dados disponível em https://github.com/nathNath/tcc_music. Acesso em 26 de junho de 2023.

confiável e verificada por um músico profissional e pela autora do trabalho, que também tem vivência na área de música. A nova base de dados disponibilizada por esta pesquisa possui uma quantidade pequena de músicas, pois o foco é trabalhar nas divergências geradas por anotações distintas entre os especialistas e os resultados das ferramentas, tanto o *framework* MSAF [1], quanto o aplicativo Moises.

## 1.2 OBJETIVOS

### 1.2.1 Objetivo Geral

Analisar e classificar as segmentações das cinco músicas mais populares do Spotify[2], com o *framework* MSAF [1], já conhecido na literatura, além do aplicativo de música Moises, utilizando uma base de dados construída pela autora deste trabalho e por um músico profissional.

### 1.2.2 Objetivos Específicos

- Construir uma base de dados que tenha uma alta confiabilidade no que diz respeito à segmentação e rotulação das cinco músicas mais populares do *Spotify*.
- Segmentar as músicas para construir a base de dados com a avaliação de um músico profissional e da autora.
- Analisar os resultados levando em consideração a eficácia da utilização de técnicas de informática musical, utilizando o *framework* de código aberto MSAF [1], e o aplicativo de música Moises, para a segmentação das cinco músicas mais populares do *Spotify*.

## 1.3 ESTRUTURA DO TRABALHO

Este trabalho é composto por seis capítulos. O primeiro capítulo introduz o tema, trazendo a problemática e os objetivos. O segundo capítulo traz o referencial teórico necessário para compreensão do trabalho realizado. O Capítulo 3 apresenta uma revisão de

[2] Plataforma *Spotify* disponível em https://open.spotify.com/ . Acesso em 01 de maio de 2023.

técnicas e metodologias de trabalhos relacionados. O Capítulo 4 contém a metodologia do trabalho, com a especificação sobre a base de dados e o seu preparo, além de descrever a configuração do sistema e os experimentos realizados. O quinto capítulo conta com os resultados alcançados nos experimentos deste trabalho com a utilização do *framework* MSAF [1] e o aplicativo de música Moises. O último capítulo conclui o trabalho com as considerações finais e os trabalhos futuros.

# 2 REFERENCIAL TEÓRICO

O presente capítulo apresenta o referencial teórico necessário para compreensão deste trabalho. A primeira seção introduz o conceito da segmentação estrutural de música, além dos métodos computacionais para realizar esta atividade e suas subatividades. São abordados os dois principais algoritmos utilizados em conjunto com o *framework* MSAF [1], que utilizam uma técnica de aprendizagem de máquina não supervisionada. É apresentada ainda outra estratégia para a resolução da problemática desta pesquisa. Por fim, são explicadas as métricas que fazem parte do processo de análise das segmentações de músicas.

## 2.1 Segmentação estrutural de música

A segmentação estrutural de música consiste em identificar segmentos de um sinal de áudio utilizando as notações "verso", "refrão", "solo", entre outros. De acordo com Nieto et. al. (2016), a segmentação estrutural de música é normalmente dividida em dois subproblemas: detecção de limites e agrupamento estrutural. O primeiro identifica o começo e o fim temporal de cada segmento da música em uma determinada parte, e o último agrupa esses segmentos baseados na sua similaridade acústica.

Esses dois subproblemas normalmente são feitos manualmente, o que demanda tempo e conhecimento por parte do músico especialista que realiza esta atividade. Pensando em abordar estes dois subproblemas da segmentação estrutural de músicas, a pesquisa de informática musical (MIR, do inglês *Music Informatics Research*) voltou-se para o desenvolvimento de métodos computacionais que realizam estas atividades. Nas subseções a seguir, serão abordados dois métodos que foram escolhidos como os principais, para a realização dos experimentos desta pesquisa, por apresentarem os resultados mais relevantes em Nieto et. al., (2016).

## 2.2 Métodos computacionais para a realização de segmentações de músicas

O ponto de partida para os métodos computacionais de segmentações de músicas envolve a extração de características de sinais de áudio puros. Estes métodos são tipicamente ajustados e otimizados empregando conjuntos de dados com anotações manuais (Nieto et. al., 2020). A seguir, serão apresentados dois métodos computacionais que serão aplicados nos

experimentos desta pesquisa: *Convex Non-Negative Matrix Factorization* (CNMF) [3] e *2D Fourier Magnitude Coefficients* (2DFMC) [8].

### 2.2.1 *Convex Non-Negative Matrix Factorization*

A abordagem de Nieto et. al. (2013) é baseada no método *Non-Negative Matrix Factorization* (NMF, do português Fatoração de Matriz Não-Negativa), que foi ampliado adicionando uma restrição convexa que resulta em centróides de agrupamento ponderados, que representam as segmentações diferentes de uma música de uma maneira mais eficiente. Esse algoritmo também consegue realizar as duas atividades da segmentação de música: encontrar os limites temporais e o agrupamento estrutural das músicas.

Para encontrar os limites temporais, Nieto et. al. (2013) executaram a técnica de agrupamento *k-means* com $k = 2$ para cada uma das matrizes CNMF de decomposição, interpretando-as como recursos de vetor de linha. Eles encontraram os limites das seções que melhor dividiram cada seção das matrizes, olhando para similaridades locais e globais devido à matriz de auto similaridade (do inglês, *Self-similarity Matrix* - SSM).

Para realizar o agrupamento estrutural das músicas, Nieto et. al. (2013) utilizaram como ideia principal as diagonais das matrizes CNMF de decomposição para compor um novo espaço de características, para que a partir dele seja possível agrupar as diferentes seções, utilizando os limites temporais encontrados na atividade anterior. Para o agrupamento, a técnica de distância Euclidiana foi utilizada. Segundo os autores, a principal desvantagem desse método é decidir o número de agrupamentos *k*, que são utilizados para agrupar novos espaços de características e é um parâmetro altamente sensível, dependendo do estilo musical do conjunto de dados. Além do algoritmo CNMF, o algoritmo 2DFMC apresentou resultados promissores em Nieto et. al., (2016)., portanto, ele também foi escolhido como objeto de análise desta pesquisa. A seguir, é feita uma breve descrição do funcionamento do algoritmo escolhido para ser comparado com o CNMF.

### 2.2.2 2D *Fourier Magnitude Coefficients*

Segundo Nieto et. al. (2014), o estado da arte, quando se trata de métodos de similaridade de segmentos de música, é o 2D *Fourier Magnitude Coefficients* (2DFMC, do português, Coeficientes de Magnitude de Fourier 2D). Esse método projeta características

harmônicas na representação 2DFMC, que permite que o método seja invariável mesmo que exista uma mudança de tom na música, ou mudança de andamento, e ele pode ser agrupado eficientemente utilizando o *k-means* para segmentar e identificar os segmentos resultantes.

Tanto o algoritmo CNMF, quanto o algoritmo 2DFMC utilizam o *k-means* [12] como técnica de agrupamento. O *k-means* funciona da seguinte forma: dado um agrupamento inicial que ainda não está otimizado, os membros do grupo são realocados para o seu centro mais próximo a cada iteração; os centros dos agrupamentos são atualizados, calculando a média dos membros de cada grupo. Isso acontece *k* vezes, em que *k* é o número de grupos a serem criados no fim da execução do algoritmo. O *k-means* é um algoritmo clássico de agrupamento.

Segundo Nieto et. al., (2014), os autores exploram vários métodos para obter um conjunto síncrono de Coeficientes de Magnitude de Fourier 2D, que podem ser utilizados para caracterizar a similaridade entre segmentos de um sinal de áudio. Isto resulta em um processo simples e pouco custoso computacionalmente. Os autores explicam que calculam o logaritmo dos trechos das músicas, para que a componente DC (frequência zero) seja relevada e as maiores frequências sejam enfatizadas. Por se tratar de um sinal de áudio simétrico, os autores exploram essa característica do sinal 2DFMC, para remover metade dos coeficientes e reduzir o custo computacional. Depois desse tratamento inicial no sinal, o *k-means* é utilizado em conjunto com a distância Euclidiana nesses trechos. Os autores explicam que, o valor de k que apresenta o melhor resultado na medida-F, é o número total de segmentações anotadas na verdade fundamental. Tendo isso em mente, será apresentado a seguir outra proposta de resolução com o uso de redes neurais convolucionais, para abordar a problemática das segmentações de música.

## 2.3 Aprendizagem de máquina profunda para segmentações de músicas

Como mencionado anteriormente, a realização da análise estrutural de música demanda tempo e conhecimento de um músico especialista. McCallum (2019) propõe o uso de uma rede neural convolucional para identificar as segmentações sem a necessidade de intervenção humana. Para cada camada convolucional da rede neural construída no trabalho de McCallum (2019), as dimensões representam o tempo, a frequência e o canal, respectivamente. Cada uma dessas camadas aplicam uma ativação ReLU. Todas as camadas convolucionais usam *kernels* 6 x 4, de acordo com a Figura 1. Depois que as características de áudio são aprendidas pela rede neural, a técnica *checkerboard kernel* é utilizada para

segmentar os limites temporais das músicas. O autor evidencia que a quantidade atual de conjuntos de dados existente na literatura é considerada baixa para o treinamento profundo de uma máquina, mas que, mesmo com a quantidade pequena desses conjuntos de dados, resultados promissores foram obtidos em sua pesquisa [13].

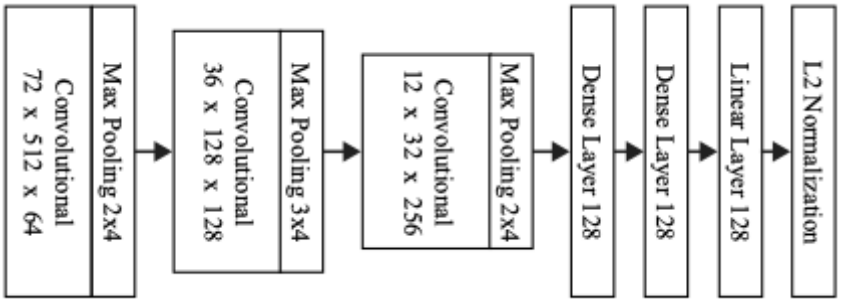



**Figura 1**: Arquitetura da rede neural utilizada nos experimentos do trabalho de McCallum. Para cada camada convolucional, as dimensões representam tempo, frequência e canal respectivamente. **Fonte**: McCallum, 2019.

Para avaliarmos a performance dos algoritmos, sendo eles supervisionados ou não, existe uma convenção presente na literatura para que seja feita a medição da precisão, *recall*, medida-F, que pondera a precisão e o *recall*, entropia e taxa de acerto, que serão apresentadas adiante.

## 2.4 Métricas de segmentações das músicas

Segundo Nieto et. al. (2016), a métrica mais padronizada para medir a qualidade do subproblema do agrupamento estrutural é a *Pairwise Frame Clustering* (do português, Agrupamento de Pares Similares de Quadros) (Tabela 1). Esta métrica compara cada par de *frames* checando se eles pertencem ao mesmo rótulo, ou seja, se pertencem ao mesmo segmento de música, tanto para a estimativa, quanto para a referência. A razão entre esses dois conjuntos de pares de *frames* sobre o número de pares similares na estimativa produzem a métrica de precisão, enquanto o *recall* é a razão entre os dois conjuntos sobre o número de pares similares de referência. A medida-F (do inglês, *F-measure*) leva em consideração essas duas métricas. Estas métricas estão definidas no MIREX[3] e implementadas no *framework* mir eval [9], que é utilizado pelo MSAF [1].

[3] Site do MIREX - Music Information Retrieval Evaluation eXchange disponível em https://www.music-ir.org/mirex/wiki/MIREX_HOME . Acesso em 21 de maio de 2023.

As métricas "*Hit Rate 3F*", "*Hit Rate 3P*", "*Hit Rate 3R*", "*Hit Rate 0.5P*", "*Hit Rate 0.5R*", "*Hit Rate 0.5F*", "*Hit Rate t3P*", "*Hit Rate t3R*", "*Hit Rate t3F*", "*Hit Rate t0.5F*", "*Hit Rate t0.5P*" e "*Hit Rate t0.5R*" foram utilizadas para medir a performance de limites temporais das músicas (carimbos de hora) (Tabela 2).

| **Métrica** | **Descrição** |
|---|---|
| *PWF* | Medida-F do *Pairwise Frame Clustering* |
| *PWP* | Precisão do *Pairwise Frame Clustering* |
| *PWR* | *Recall* do *Pairwise Frame Clustering* |
| *Sf* | Pontuação da Entropia Normalizada da Medida-F |
| *So* | Pontuação da Entropia Normalizada da Precisão |
| *Su* | Pontuação da Entropia Normalizada do *Recall* |

**Tabela 1**: Métricas utilizadas para a avaliação de recuperação de informações musicais, especificamente de segmentações de músicas e suas respectivas descrições. **Fonte**: Nieto (2016). Disponível em https://msaf.readthedocs.io/en/latest/eval.html.

| **Métrica** | **Descrição** |
|---|---|
| *Hit Rate 3F* | Taxa de acerto da medida-F usando uma janela de 3 segundos |
| *Hit Rate 3P* | Taxa de precisão usando uma janela de 3 segundos |
| *Hit Rate 3R* | Taxa de acerto do *recall* usando uma janela de 3 segundos |
| *Hit Rate 0.5P* | Taxa de precisão usando uma janela de meio segundo |
| *Hit Rate 0.5R* | Taxa de acerto do *recall* usando uma janela de meio segundo |
| *Hit Rate 0.5F* | Taxa de acerto da medida-F usando uma janela de meio segundo |
| *Hit Rate t3P* | Taxa de precisão usando a janela de 3 segundos sem o primeiro e último limites da música |

| | |
|---|---|
| *Hit Rate t3R* | Taxa de *recall* usando a janela de 3 segundos sem o primeiro e último limites da música |
| *Hit Rate t3F* | Taxa de acerto da medida-F usando a janela de 3 segundos sem o primeiro e último limites da música |
| *Hit Rate t0.5F* | Taxa de acerto da medida-F usando a janela de 0,5 segundo sem o primeiro e último limites da música |
| *Hit Rate t0.5P* | Taxa de precisão usando a janela de 0,5 segundo sem o primeiro e último limites da música |
| *Hit Rate t0.5R* | Taxa de *recall* usando a janela de 0,5 segundo sem o primeiro e último limites da música |

**Tabela 2**: Métricas utilizadas para a avaliação de recuperação de informações musicais, especificamente de limites temporais das músicas. **Fonte**: Nieto (2016). Disponível em https://msaf.readthedocs.io/en/latest/eval.html.

Segundo Davies et. al. (2009), A medida-F é uma métrica genérica para avaliação de recuperação de informações musicais. A medida-F é calculada levando em consideração três parâmetros: $c$, o número de detecções corretas (verdadeiros positivos), $f^-$, o número de falsos negativos (detecções perdidas) e $f^+$, o número de falsos positivos (detecções extras). Uma detecção correta é considerada levando em conta uma janela de tolerância ao redor de uma determinada anotação. A medida-F é calculada usando duas quantidades intermediárias, precisão, $p$ e *recall*, $r$. A precisão indica a proporção das predições geradas que estão corretas,

$$p = \frac{c}{c + f^+} \tag{1}$$

e o *recall* indica a proporção do número total de anotações corretas que foram encontradas,

$$r = \frac{c}{c + f^-} . \tag{2}$$

Quando combinadas, elas fornecem o valor de precisão medida-F,

$$F = \frac{2pr}{p+r} = \frac{2c}{2c+f^{+}+f^{-}} \times 100\% . \quad (3)$$

# 3 TRABALHOS RELACIONADOS

Este capítulo visa apresentar trabalhos relacionados à proposta desta pesquisa, e mostra de uma forma mais sucinta quais são as técnicas e as metodologias que foram aplicadas para atingir os resultados destes trabalhos, além dos objetivos, número de músicas, número de estilos musicais e a quantidade de avaliadores especialistas envolvidos no trabalho.

O trabalho intitulado "*Evaluation Methods for Musical Audio Beat Tracking Algorithms*" [5] tem como objetivo propor uma nova técnica para extrair as localizações das batidas de músicas de forma automática. O presente trabalho se relaciona de alguma forma com a presente proposta, pois o *framework* utilizado nos experimentos deste trabalho também conta com uma funcionalidade que consegue localizar de forma automática as batidas das músicas. A principal relação entre os dois trabalhos é a criação de um novo conjunto de dados para testes que serão disponibilizados para a comunidade científica. O conjunto de dados utilizado no trabalho relacionado consiste na anotação das batidas de 179 músicas da banda de rock *The Beatles*, popularmente conhecida no mundo. O conjunto de dados do trabalho relacionado foi curado por dois músicos especialistas, segundo os autores.

Em "*Design and Creation of a Large-scale Database of Structural Annotations*" [6], os autores disponibilizaram para a comunidade um grande conjunto de dados com mais de 2400 anotações das segmentações de músicas e quase 1400 gravações musicais. O trabalho deles demonstrou os objetivos da elaboração deste conjunto de dados, bem como os desafios que eles encontraram para realizar a atividade em si. Segundo os autores, os mesmos enfrentaram muitos problemas metodológicos com uma grande base de dados, devido à escolha dos estilos de música variados, formato das anotações das segmentações e procedimentos. Também pontuaram o alto nível de subjetividade encontrado para realizar a rotulação das segmentações de músicas, o que contribui para um alto nível de discordância entre os avaliadores das músicas.

Em relação a este trabalho, compartilhamos o objetivo de criar mais conjuntos de dados para testes que sirvam à comunidade científica. Eles focaram em quatro estilos musicais e contrataram oito músicos em formação nas áreas de Teoria Musical ou Composição. Eles tiveram como objetivo calcular o tempo gasto para realizar as anotações das segmentações de músicas e também calcularam a similaridade das anotações das músicas, entre os contratados, para realizar a atividade no escopo da pesquisa.

Em "*JAMS: A JSON Annotated Music Specification for Reproducible MIR Research*" [7], os autores elaboraram um novo modelo de especificação para anotação de músicas voltadas para a pesquisa científica, na área de recuperação de informações de música. Eles visam a simplicidade, estrutura e sustentabilidade dessas anotações. Também tiveram como objetivo a transcrição e a disponibilização de conjuntos de dados famosos na literatura para essa nova notação proposta, a fim de facilitar que os pesquisadores da área comecem a utilizar a nova padronização. A relação entre o presente trabalho e o JAMS é direta, haja vista que o *framework* MSAF funciona com a utilização de um arquivo JAMS de entrada. Para a execução do experimento deste trabalho foi necessário converter as anotações, inicialmente feitas em formato textual, para depois serem convertidas em formato CSV, para enfim, adotarem o formato "JAMS", que é uma estrutura elaborada de arquivo em formato JSON.

Na Tabela 3, são apresentados os títulos dos trabalhos relacionados, quais são os objetivos, número de músicas, número de estilos musicais e a quantidade de avaliadores especialistas envolvidos no trabalho (Tabela 3).

| **Título do trabalho** | **Objetivo** | **Número de músicas** | **Estilos musicais** | **Quantidade de avaliadores especialistas** |
|---|---|---|---|---|
| "*Evaluation Methods for Musical Audio Beat Tracking Algorithms*" | Propor uma nova técnica para extrair as localizações das batidas de músicas de forma automática. | 179 músicas. | Rock britânico. | Dois. |
| "*Design and Creation of a Large-scale Database of Structural Annotations*" | Disponibilizar para a comunidade um grande conjunto de dados com mais de 2400 anotações das | 2400 músicas. | Pop, Jazz, Classical, World Music. | Oito. |

| | | | | |
|---|---|---|---|---|
| | segmentações de músicas e quase 1400 gravações musicais. | | | |
| "*JAMS: A JSON Annotated Music Specification for Reproducible MIR Research*" | Elaborar um novo modelo de especificação para anotação de músicas voltadas para a pesquisa científica na área de pesquisa informática musical. | Não informado. | Rock, Pop, Opera. | Não informado. |
| "Análise e classificação das segmentações de músicas " | Analisar e classificar as segmentações das cinco músicas mais populares do Spotify, com o *framework* MSAF, já conhecido na literatura, além do aplicativo de música Moises, utilizando uma base de dados construída pela | 5 músicas. | *Pop*, *K-pop*, *Reggaeton*, *Alternative/Indie*, *Latin Urbano*. | Dois. |

| | | | | |
|---|---|---|---|---|
| | autora deste trabalho e por um músico profissional. | | | |

**Tabela 3**: Trabalhos relacionados com os seus respectivos títulos, objetivos, número de músicas, número de estilos musicais e quantidade de avaliadores especialistas envolvidos no trabalho. **Fonte**: Autora, 2023.

# 4 METODOLOGIA

Este trabalho tem o intuito de contribuir com o campo da pesquisa de informática musical, que possui como um de seus principais desafios a variedade e riqueza de anotações de segmentações de músicas que sejam produzidas por pessoas confiáveis e com conhecimento teórico e prático em música. Serão discorridos, nos tópicos a seguir, os passos a serem realizados para construir a base de dados, as técnicas que serão utilizadas na preparação da base de dados e a configuração do sistema que será utilizada neste projeto, além de especificar como será conduzido o experimento desta pesquisa.

## 4.1 BASE DE DADOS

Um dos objetivos deste trabalho é produzir uma base de dados que seja considerada padrão-ouro, verificada por um músico profissional, atuante no mercado de trabalho e na academia. As músicas selecionadas são acessíveis ao público que possui acesso à Internet, para suas segmentações por um especialista, sendo utilizada, dessa forma, para avaliação de modelos de segmentação musical automática disponíveis na literatura.

Serão disponibilizadas as classificações de segmentações das músicas "Cupid" - Fifty Fifty, "Ella Baila Sola" - Eslabon Armado e Peso Pluma, "Flowers" - Miley Cyrus, "La Bebe" - Remix - Yng Lvcas e Peso Pluma e "Unx100to" - Grupo Frontera e Bad Bunny, que estão na lista das músicas mais populares do mundo da plataforma de *streaming* Spotify[4], que pode ser acessada gratuitamente por qualquer pessoa que possua uma conta nela. A base de dados deste trabalho também será disponibilizada de forma pública e gratuita na plataforma Github, a fim de contribuir com a comunidade científica que deseje utilizar a base de dados para trabalhos futuros.

Todas as classificações de segmentações de músicas serão feitas de forma manual e independente, com a atuação da autora do trabalho, ex-estudante de teoria musical da Escola de Música Anthenor Navarro (EMAN), situada em João Pessoa/PB, atual estudante de teoria e prática musical com foco em guitarra elétrica com o professor Gustavo Queiroga. Contará também com as classificações de Gustavo Queiroga, produtor musical e músico profissional, que é graduado em Música Popular pela UFPB, com especialização em guitarra elétrica.

[4] Plataforma Spotify disponível em https://open.spotify.com/ . Acesso em 01 de maio de 2023.

Nesse tempo trabalhou com importantes artistas do cenário musical brasileiro, tais como Felipe Andreoli (Angra), Bruno Valverde (Angra), Nando Fernandes, Renata Arruda, entre outros. Atualmente tem o seu projeto solo intitulado GQ Project e lidera a banda de *heavy metal* Antimatter Life. As classificações da autora e do músico especialista foram feitas de forma separada, para que não houvesse enviesamento das classificações das segmentações.

O processo aconteceu da seguinte forma: a pessoa classificadora ouve a música duas vezes; na primeira vez são feitas anotações rápidas das seções das músicas, que podem ser classificadas geralmente, mas não exclusivamente em "*intro, verse, pre-chorus, chorus, solo, bridge, silence*". Na segunda vez, foram feitos ajustes finos nas marcações de tempo, e até mesmo reajustes nas rotulações das segmentações, caso seja notado algum novo detalhe na segunda audição da música. A autora do trabalho optou por rotular as seções das músicas em inglês, a fim de atingir um público maior com a disponibilização da base de dados posteriormente. Na Tabela 4, é demonstrado um exemplo de rotulação de música. Ao todo, foram produzidos cinco arquivos similares ao modelo apresentado nesta tabela, em formato de arquivo CSV.

| **Título da música e intérprete** | **Avaliadora** | **Rotulação da música** |
|---|---|---|
| Flowers - Miley Cyrus | Autora | 0:00 - introdução<br>0:08 - verso<br>0:24 - pré-refrão<br>0:33 - refrão<br>1:10 - verso<br>1:26 - pré-refrão<br>1:33 - refrão<br>2:19 - pré-refrão<br>2:27 - refrão<br>2:52 - pós-refrão<br>3:00 - outro<br>3:16 - silêncio |

**Tabela 4**: Rotulação da música intitulada Flowers, interpretada por Miley Cyrus, avaliada e rotulada pela autora deste trabalho. **Fonte**: Autora, 2023.

## 4.2 PREPARAÇÃO DA BASE DE DADOS

Após a criação da base de dados padrão-ouro, é necessário buscar e possuir os arquivos em formato MP3 das cinco músicas mais populares do mundo da plataforma de *streaming* Spotify. Para possuir estes arquivos de forma legal, foi utilizada a plataforma *TuneFab*[5], com licença experimental (*trial*) em conjunto com uma conta *premium* da própria autora, da plataforma Spotify. Vale ressaltar que estes arquivos extraídos em formato MP3 não são disponibilizados para o público, a fim de não ferir nenhum tipo de direito autoral. Os arquivos extraídos são utilizados única e exclusivamente para uso pessoal, com o intuito de realizar pesquisa científica, que não está vinculada a nenhum tipo de produto ou interesse comercial. Com os arquivos das músicas no formato MP3 em mãos, conseguiremos preparar a base de dados a ser construída neste trabalho, para que ela sirva de referência ao realizar as avaliações das músicas.

## 4.3 CONFIGURAÇÃO DO SISTEMA

O sistema está configurado para operar com a linguagem interpretada Python, versão 3.10.8. Foi utilizado o Visual Studio Code como IDE para este projeto. O *framework* de código aberto e livre chamado MSAF [1] é o principal a ser utilizado nesta pesquisa. Segundo Nieto (2016), o MSAF [1] é um pacote em Python para a análise de algoritmos de segmentação estrutural de músicas. Ele possui um conjunto de recursos, algoritmos, métricas de avaliação e conjuntos de dados para experimentação. Ao todo, existem oito algoritmos implementados, que foram feitos levando em consideração outras referências com código aberto, ou até mesmo algoritmos feitos do zero, por não possuírem código aberto. Eles têm a função de fazer a classificação dos limites entre as seções das músicas e agrupar as estruturas musicais. O aplicativo de música Moises, Versão 1.0.32 para MacOS, também foi utilizado no trabalho para que a sua análise de seções fosse extraída e comparada com as anotações consideradas verdades fundamentais.

[5] Plataforma TuneFab disponível em https://www.tunefab.com/ . Acesso em 01 de maio de 2023.

## 4.4 EXPERIMENTO

O experimento consiste em utilizar o *framework* MSAF [1] e o aplicativo Moises para realizar a detecção dos limites das músicas, bem como rotular as segmentações e comparar os resultados obtidos pelos algoritmos com as referências criadas pela autora e pelo músico profissional. Foi dada a preferência para algoritmos que já estão implementados no *framework* MSAF e que têm a possibilidade de detectar os limites e fazer os agrupamentos das estruturas, como, por exemplo, *Constrained Cluster* [2], *Convex NMF* [3] e *Laplacian Segmentation* [4]. Ao todo, foram realizados 3 experimentos, com pelo menos 5 músicas cada, utilizando o MSAF [1], o aplicativo de música Moises e as anotações consideradas verdades fundamentais do músico profissional e da autora do trabalho.

Para os experimentos deste trabalho, foram utilizadas as mesmas configurações do trabalho MSAF [1], consideradas padrão para o *framework*: a taxa de amostragem de 11kHz, FFT e os tamanhos dos *hop sizes* foram 2048 e 512 amostras, respectivamente. O tipo da funcionalidade padrão foi o PCP (representa a harmonia como característica principal); número de oitavas e frequência inicial para os PCPs foram 7 e 27.5Hz respectivamente; as métricas de avaliação foram a "*Hit Rate 3F*" e o *Pairwise Frame Clustering* para a detecção dos limites e agrupamento estrutural das músicas, respectivamente (Nieto et. al., 2016).

Além da análise da segmentação de músicas com o *framework* MSAF [1], também foi feita uma análise com as segmentações feitas pelo aplicativo de música Moises[6], habilitando a funcionalidade "Seções - Beta", que não possui nenhum tipo de parametrização para os usuários finais, que separa as seções das músicas utilizando um algoritmo de inteligência artificial, o qual não foi divulgado abertamente por sua página oficial, por se tratar de um produto comercializável no mercado. Entretanto, por ser um aplicativo de música que vem ganhando relevância no meio musical e de aplicativos, ele também foi escolhido como objeto de análise do experimento.

[6] Site do Moises disponível em https://studio.moises.ai/ . Acesso em 29 de maio de 2023.

# 5 RESULTADOS

Tanto o experimento Fourier contra as anotações do músico profissional, quanto o experimento CNMF comparado com as anotações do músico profissional, utilizaram o algoritmo *Structural Features* [10] para metrificar os limites de tempo das músicas, a fim de fixar os carimbos de hora, para que a performance dos algoritmos de segmentação de música pudesse ser comparada adequadamente. A principal diferença entre os dois experimentos é a aplicação de algoritmos de segmentação de música distintos, onde o experimento A utilizou o algoritmo *2D-Fourier Magnitude Coefficients* [8], e o experimento B utilizou o algoritmo *Convex Non-Negative Matrix Factorization* [3].

Na literatura, o valor de referência adotado para mensurar a taxa de acerto em relação aos limites das músicas é a "*Hit Rate* 3F", que leva em consideração a medida-F em uma janela de 3 segundos. A *Hit Rate 3F* da música Cupid foi de 72%, ou seja, o algoritmo utilizado para metrificar os limites temporais das músicas alcançou um resultado considerado satisfatório ao definir os limites das músicas, comparado às anotações do músico especialista.

Para a música Ella Baila Sola, consta a "*Hit Rate 3F*" de 35,3%, que é a taxa de acerto levando consideração a medida-F em uma janela de 3 segundos, ou seja, o algoritmo de detecção de limites da música não apresentou um resultado satisfatório, em relação às anotações do músico especialista.

Para a música Flowers de Miley Cyrus, a taxa de acerto do algoritmo de classificação de limites das músicas em relação às anotações do músico especialista foi de apenas 31,57%. A taxa de acerto apresentou o resultado mais satisfatório dos experimentos, com o valor de 78,57%, para a música La Bebe, dos artistas Yng Lvcas e Peso Pluma. Por fim, a taxa de acerto foi de apenas 25% para a música unx100to, dos artistas Grupo Frontera e Bad Bunny. Esta taxa de acerto dos limites temporais da música foi a menor de todas as que foram realizadas nos experimentos, sempre utilizando as anotações do músico especialista como referência (Figura 2).

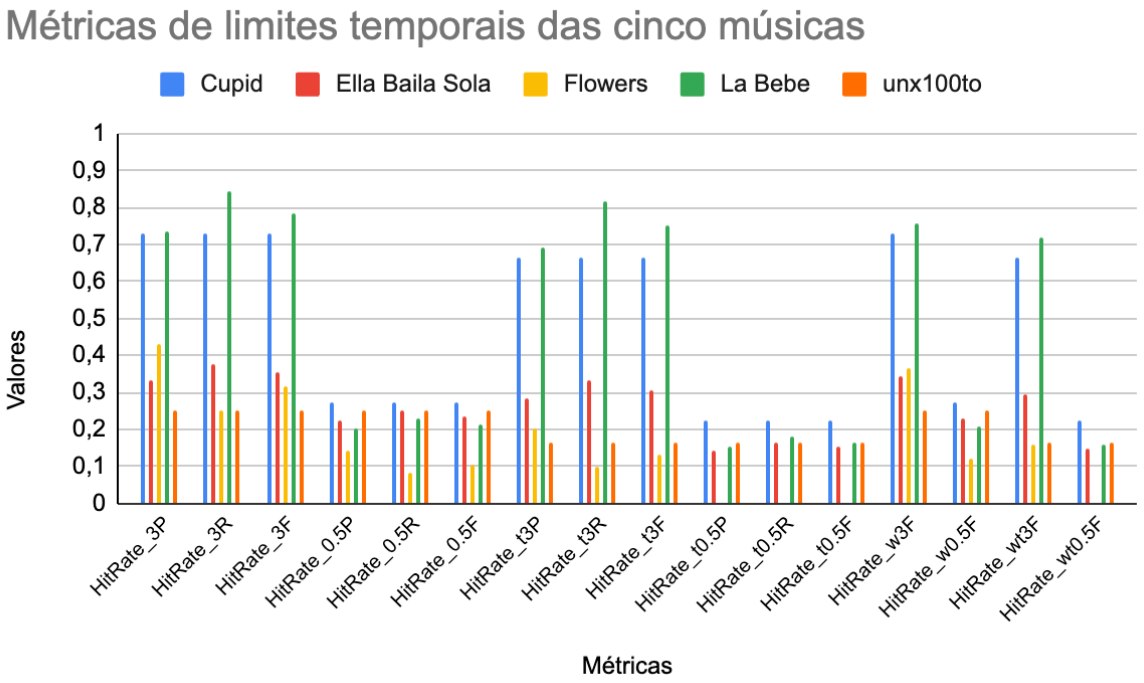



**Figura 2**: Resultados das métricas de limites temporais das cinco músicas nos experimentos A e B.

**Fonte**: Autora, 2023.

## 5.1 EXPERIMENTO A: MSAF FOURIER COMPARADO COM AS ANOTAÇÕES DO MÚSICO PROFISSIONAL

O Experimento A consistiu em utilizar as configurações padrão, com o algoritmo de segmentação de música *2D-Fourier Magnitude Coefficients* [8]. Em relação à segmentação das músicas, as principais métricas analisadas foram a Precisão do *Pairwise Frame Clustering*, *Recall* do *Pairwise Frame Clustering*, Medida-F do *Pairwise Frame Clustering*, Pontuação da Entropia Normalizada da Precisão, Pontuação da Entropia Normalizada do *Recall* e a Pontuação da Entropia Normalizada da Medida-F (Figura 3).

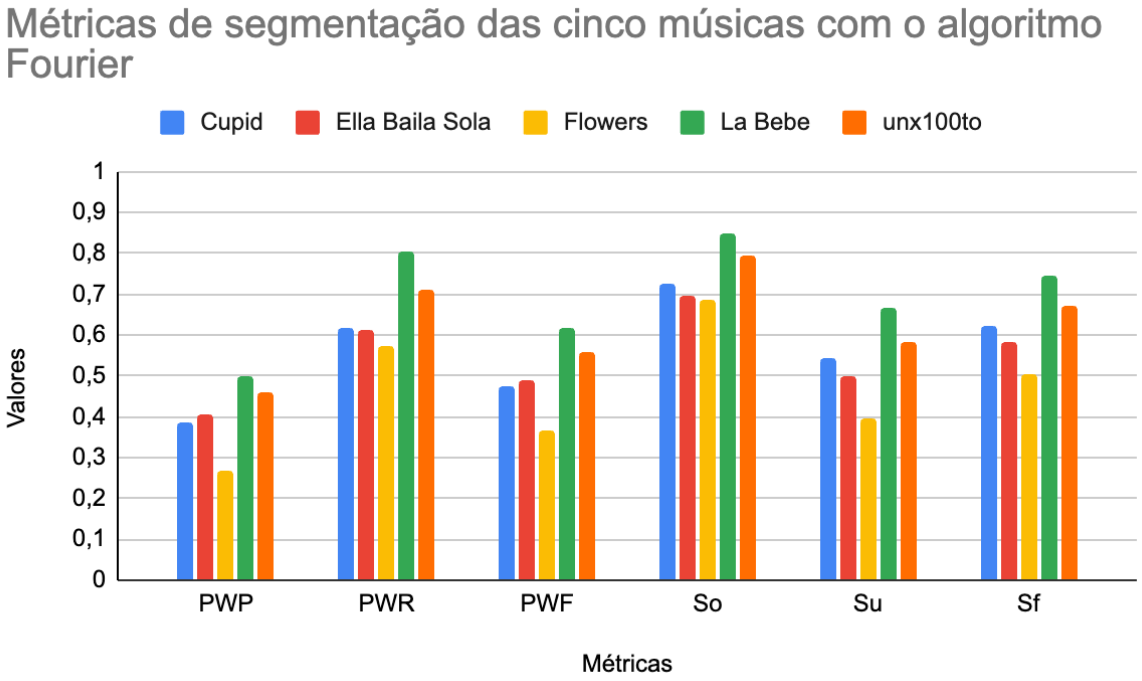



**Figura 3:** Resultados das métricas precisão do *Pairwise Frame Clustering*, *recall* do *Pairwise Frame Clustering*, medida-F do Pairwise Frame Clustering, pontuação da entropia normalizada da precisão, pontuação da entropia normalizada do *recall* e a pontuação da entropia normalizada da medida-F de segmentação das cinco músicas no experimento A. **Fonte**: Autora, 2023.

Na Figura 3, é possível notar que a música La Bebe apresentou os melhores resultados em todas as métricas, alcançando 50,02% na métrica PWP, a precisão do *Pairwise Frame Clustering*, 80,43% na métrica PWR, o recall do *Pairwise Frame Clustering*, 61,68% na métrica PWF, a ponderação entre a precisão e o *recall* do *Pairwise Frame Clustering*, 84,96% na métrica $So$, Pontuação da Entropia Normalizada da Precisão, 66,48% na métrica $Su$, Pontuação da Entropia Normalizada do *Recall* e 74,59% na métrica $Sf$, Pontuação da Entropia Normalizada da Medida-F. Já a música Flowers apresentou os piores resultados nas métricas PWP - 26,77%, PWR - 57,46%, PWF - 36,52%, So - 68,73%, Su - 39,54% e Sf - 50,20%.

### 5.1.1 Música Cupid

Vale ressaltar que o algoritmo e o músico especialista utilizaram notações diferentes para segmentar a música. A taxa de acerto alcançada foi de 47,51% (PWF). O algoritmo *2D-Fourier Magnitude Coefficients* [8] segmentou a música intitulada Cupid do artista Fifty

Fifty em oito partes, enquanto o músico especialista dividiu a música em dez partes (Figura 4).

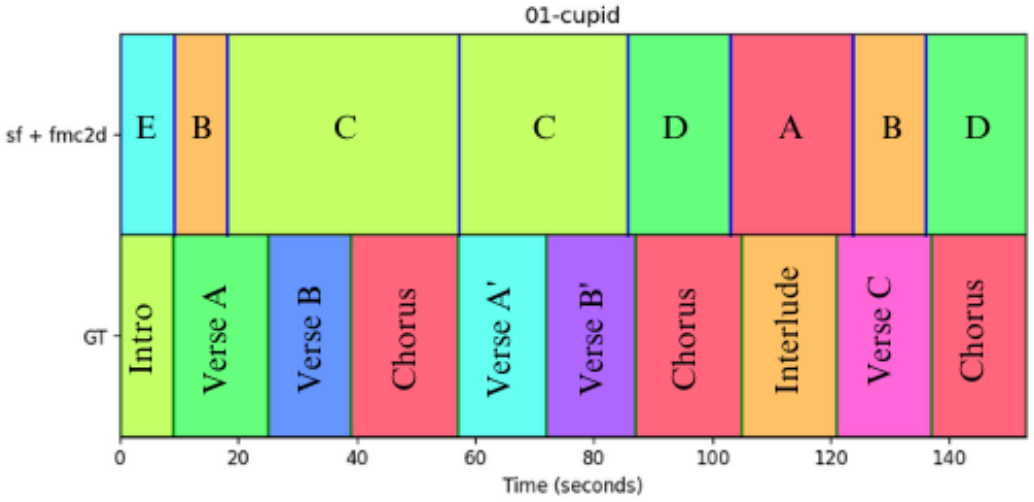



**Figura 4:** Comparação entre as segmentações feitas pelo algoritmo *2D-Fourier Magnitude Coefficients* [8], indicada por "sf+fmc2d" e as anotações feitas pelo músico especialista, indicadas por "GT", no decorrer da duração da música Cupid em segundos, no experimento A. **Fonte**: Autora, 2023.

### 5.1.2 Música Ella Baila Sola

A taxa de acerto alcançada neste experimento foi de 48,76% (PWF). O algoritmo *2D-Fourier Magnitude Coefficients* [8] segmentou a música intitulada Ella Baila Sola em seis partes, enquanto o músico especialista dividiu a música em sete partes (Figura 5).

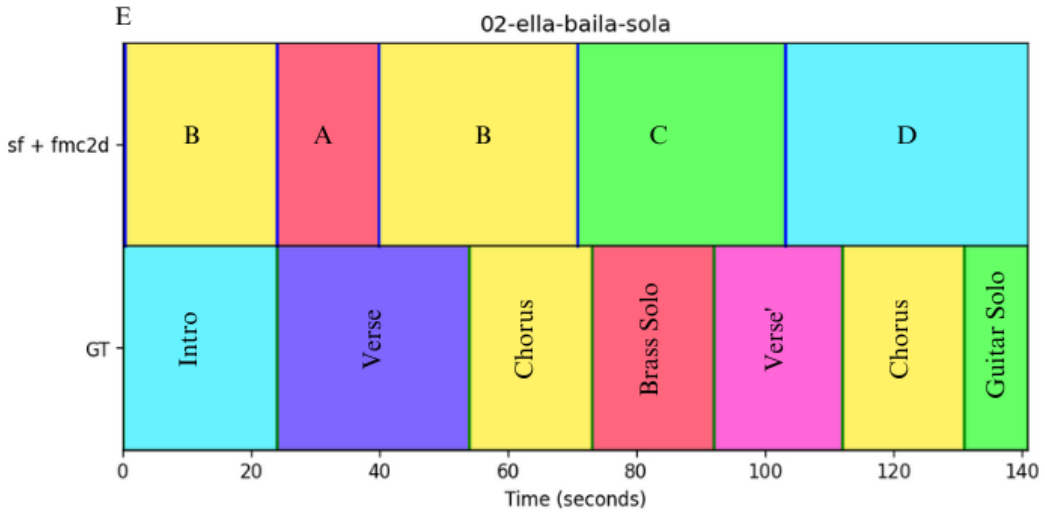


**Figura 5:** Comparação entre as segmentações feitas pelo algoritmo *2D-Fourier Magnitude Coefficients* [8], indicada por "sf+fmc2d" e as anotações feitas pelo músico especialista, indicadas por "GT", no decorrer da duração da música Ella Baila Sola em segundos, no experimento A. **Fonte**: Autora, 2023.

### 5.1.3 Música Flowers

A taxa de acerto alcançada neste experimento foi de apenas 36,52% (PWF), o pior resultado do experimento em questão. O algoritmo *2D-Fourier Magnitude Coefficients* [8] segmentou a música intitulada Flowers em cinco partes, enquanto o músico especialista dividiu a música em onze partes (Figura 6).


**Figura 6:** Comparação entre as segmentações feitas pelo algoritmo *2D-Fourier Magnitude Coefficients* [8], indicada por "sf+fmc2d" e as anotações feitas pelo músico especialista, indicadas por "GT", no decorrer da duração da música Flowers em segundos, no experimento A. **Fonte**: Autora, 2023.

### 5.1.4 Música La Bebe

A taxa de acerto alcançada neste experimento foi de 61,68% (PWF), o melhor resultado do experimento em questão. O algoritmo *2D-Fourier Magnitude Coefficients* [8] segmentou a música intitulada La Bebe em quatorze partes, enquanto o músico especialista dividiu a música em doze partes (Figura 7).

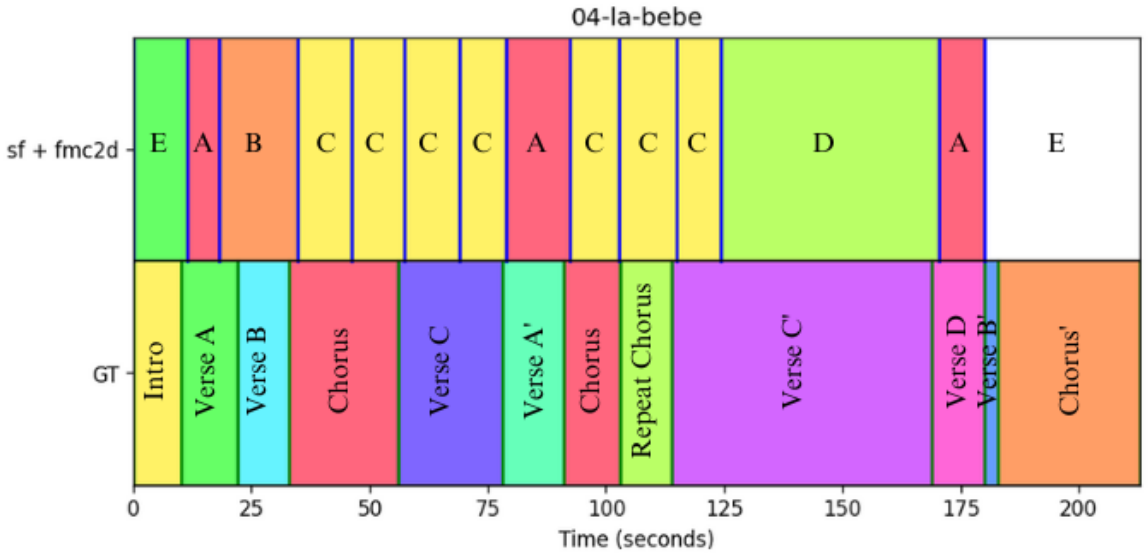



**Figura 7:** Comparação entre as segmentações feitas pelo algoritmo *2D-Fourier Magnitude Coefficients* [8], indicada por "sf+fmc2d" e as anotações feitas pelo músico especialista, indicadas por "GT", no decorrer da duração da música La Bebe em segundos, no experimento A. **Fonte**: Autora, 2023.

### 5.1.5 Música unx100to

A taxa de acerto alcançada neste experimento foi de 55,95% (PWF). O algoritmo *2D-Fourier Magnitude Coefficients* [8] segmentou a música intitulada La Bebe em seis partes, enquanto o músico especialista dividiu a música em sete partes (Figura 8).

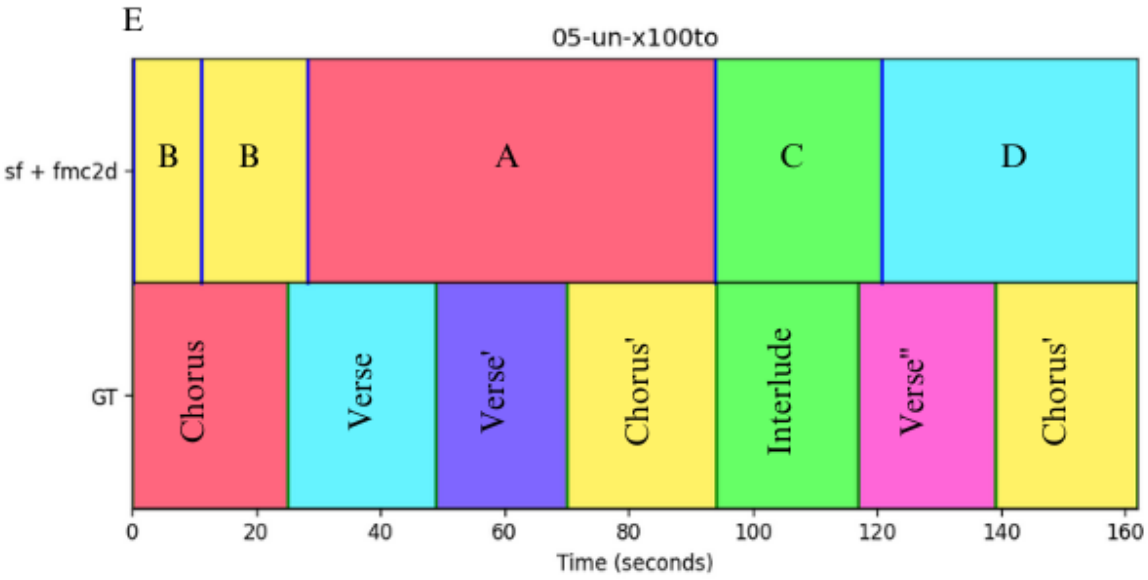



**Figura 8:** Comparação entre as segmentações feitas pelo algoritmo *2D-Fourier Magnitude Coefficients* [8], indicada por "sf+fmc2d" e as anotações feitas pelo músico especialista, indicadas por "GT", no decorrer da duração da música unx100to em segundos, no experimento A. **Fonte**: Autora, 2023.

Em Cupid, uma música do estilo *k-pop*, os algoritmos *Structural Features* (avaliação dos carimbos de hora) e 2DFMC (segmentação de músicas - criação dos rótulos) acertaram totalmente a introdução e um trecho do verso A (verdade fundamental) da música. Aos 16 segundos da música, acontece novamente uma progressão dos acordes Em7 (Mi menor com sétima), A7 (Lá maior com sétima), D (Ré maior), Bm7 (Si menor com sétima), o que pode ter induzido o algoritmo a avaliar que o segmento B (algoritmo) encerrou depois dessa progressão ter ocorrido uma vez. O verso B (verdade fundamental) é anotado aos 25 segundos e em seguida acontece o refrão, mas o algoritmo apresentou subsegmentações até o final do refrão (verdade fundamental). A segmentação não deve ter sido realizada pelo algoritmo, pois tanto a batida, quanto a melodia desse trecho da música, são bem parecidas, com leves variações no acorde D (Ré maior), que precedem o refrão, tendo o Dmaj7 (Ré maior com sétima) e o D7 (Ré com sétima). Como essa progressão de acordes acontece várias vezes no decorrer da música, isso caracteriza a **repetição**, o que pode levar o algoritmo a agrupar esses *frames* em uma seção específica. Após 1min25s de música, ela possui seções bem diferentes, caracterizando outro aspecto avaliado no algoritmo *Structural Features* [10], a **novidade**, o que contribuiu para que o algoritmo identificasse essas variações, com erros apenas na marcação de tempo. Para a interpretação das demais músicas, estas características de **novidade** e **repetição** são válidas.

Levando em consideração a métrica PWF, que pondera a precisão e o *recall*, a maior diferença de avaliação entre o algoritmo e o músico profissional aconteceu na música Flowers. A menor diferença de avaliação ocorreu na música La Bebe. Nela, a introdução é bem nítida, pois possui apenas um sintetizador marcando os acordes Gm (Sol menor) e Dm (Ré menor). Em seguida, começa uma nova seção com o começo da letra da música, com os mesmos acordes e sem percussão, o que gera o fator **novidade** para o algoritmo. Próximo aos 32 segundos, é introduzida a *drum beat* que ocorre basicamente na música toda, demarcando uma nova seção C (algoritmo). Em meio às seções C (algoritmo), existe um verso sem esse *drum beat*, que diferencia e gera uma seção A, que não possui percussão e os acordes são os mesmos. Depois, o *drum beat* é reintroduzido e o algoritmo volta a anotar a seção C aos 1min35s. Entre 2min2s e 2min5s acontece uma virada na *drum beat*, o que pode ter induzido o algoritmo a marcar uma nova seção D. Depois dessa seção, um verso sem *drum beat* volta a ocorrer, o que caracteriza a repetição da seção A. Aos 3min, o baixo da música começa a marcar os acordes de maneira diferente, o que caracteriza uma nova seção E. Em relação ao músico, o algoritmo não conseguiu identificar os refrões da música.

## 5.2 EXPERIMENTO B: MSAF CNMF COMPARADO COM AS ANOTAÇÕES DO MÚSICO PROFISSIONAL

O Experimento B consistiu em utilizar as configurações padrão, com o algoritmo de segmentação de músicas *Convex Non-Negative Matrix Factorization* (CNMF) [3]. Em relação à segmentação das músicas, as principais métricas analisadas estão representadas na Figura 9.

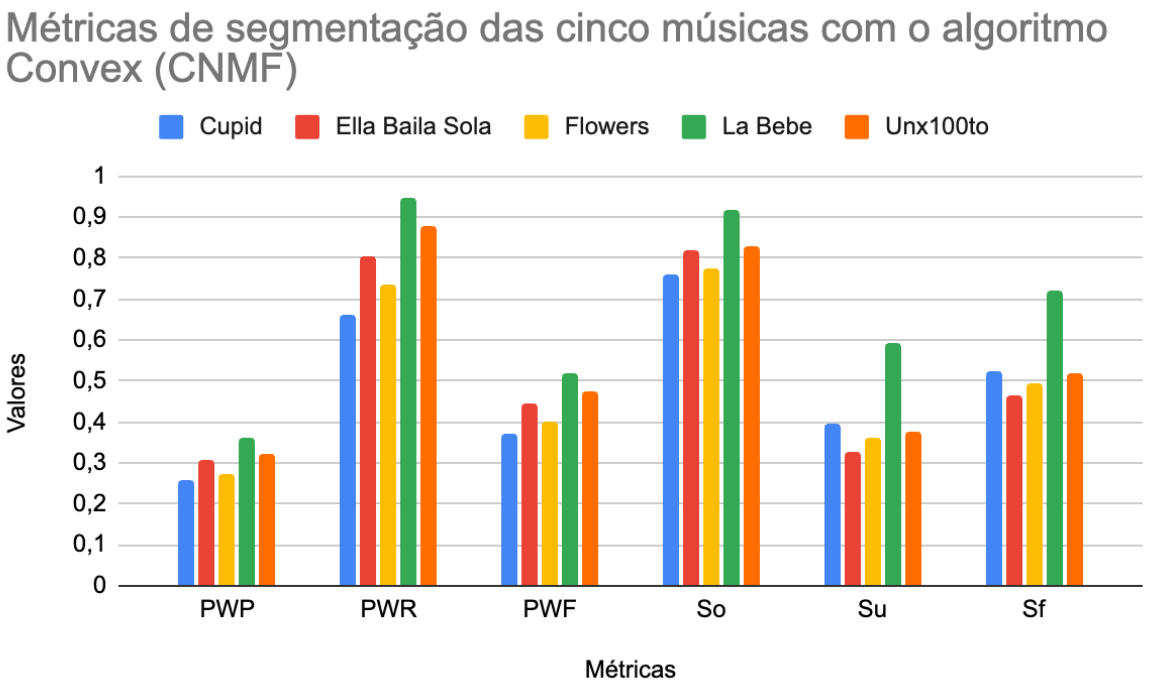



**Figura 9:** Resultados das métricas precisão do *Pairwise Frame Clustering*, *recall* do *Pairwise Frame Clustering*, medida-F do Pairwise Frame Clustering, pontuação da entropia normalizada da precisão, pontuação da entropia normalizada do *recall* e a pontuação da entropia normalizada da medida-F de segmentação das cinco músicas no experimento B. **Fonte**: Autora, 2023.

Na Figura 9, é possível notar que a música La Bebe apresentou os melhores resultados em todas as métricas, alcançando 35,92% na métrica PWP, a precisão do *Pairwise Frame Clustering*, 94,93% na métrica PWR, o recall do *Pairwise Frame Clustering*, 52,12% na métrica PWF, a ponderação entre a precisão e o *recall* do *Pairwise Frame Clustering*, 91,94% na métrica $So$, Pontuação da Entropia Normalizada da Precisão, 59,53% na métrica $Su$, Pontuação da Entropia Normalizada do *Recall* e 72,27% na métrica $Sf$, Pontuação da Entropia Normalizada da Medida-F. Diferente do algoritmo 2DFMC, a música Cupid apresentou os piores resultados nas métricas PWP - 25,60%, PWR - 66,21%, PWF - 36,92%,

So - 76,02%. A música Ella Baila Sola ficou com o pior resultado nas métricas Su - 32,55% e Sf - 46,61%.

### 5.2.1 Música Cupid

Vale ressaltar que o algoritmo e o músico especialista utilizaram notações diferentes para segmentar a música, logo, justifica a baixa taxa de acerto, que ficou em 36,92% (PWF). O algoritmo CNMF [3] segmentou a música intitulada Cupid do grupo Fifty Fifty em oito partes, enquanto o músico especialista dividiu a música em dez partes (Figura 10).

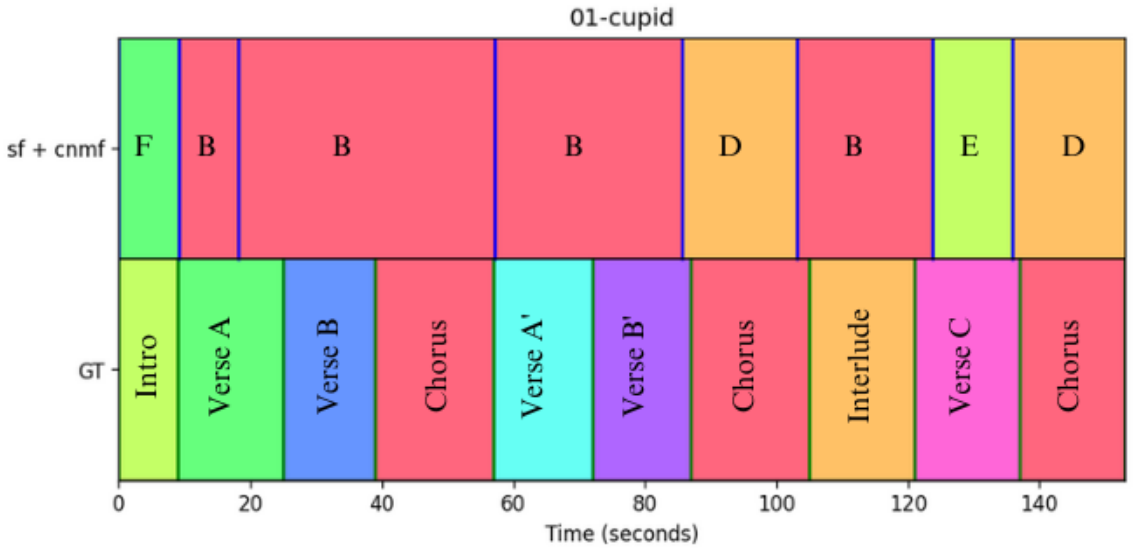



**Figura 10:** Comparação entre as segmentações feitas pelo algoritmo *Convex Non-Negative Matrix Factorization* [3], indicada por "sf+cnmf" e as anotações feitas pelo músico especialista, indicadas por "GT", no decorrer da duração da música Cupid em segundos, no experimento B. **Fonte**: Autora, 2023.

### 5.2.2 Música Ella Baila Sola

A taxa de acerto atingida na música Ella Baila Sola foi de 44,53% (PWF). O algoritmo CNMF [3] dividiu a música em apenas cinco partes, enquanto o músico especialista realizou a divisão dela em sete partes (Figura 11).

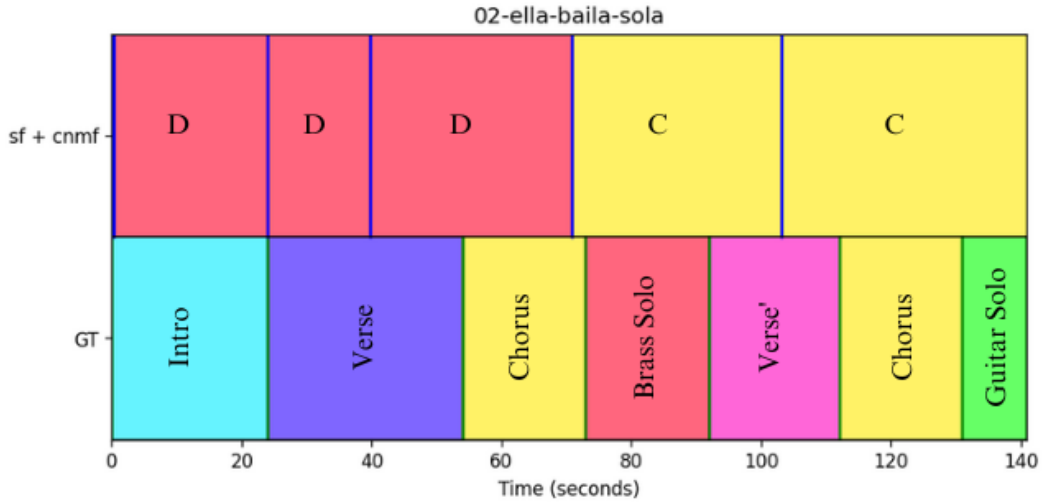



**Figura 11:** Comparação entre as segmentações feitas pelo algoritmo *Convex Non-Negative Matrix Factorization* [3], indicada por "sf+cnmf" e as anotações feitas pelo músico especialista, indicadas por "GT", no decorrer da duração da música Ella Baila Sola em segundos, no experimento B. **Fonte**: Autora, 2023.

### 5.2.3 Música Flowers

A taxa de acerto atingida na música Flowers foi de 39,97% (PWF). O algoritmo CNMF [3] dividiu a música em apenas cinco partes, enquanto o músico especialista realizou a divisão da música em onze partes (Figura 12).

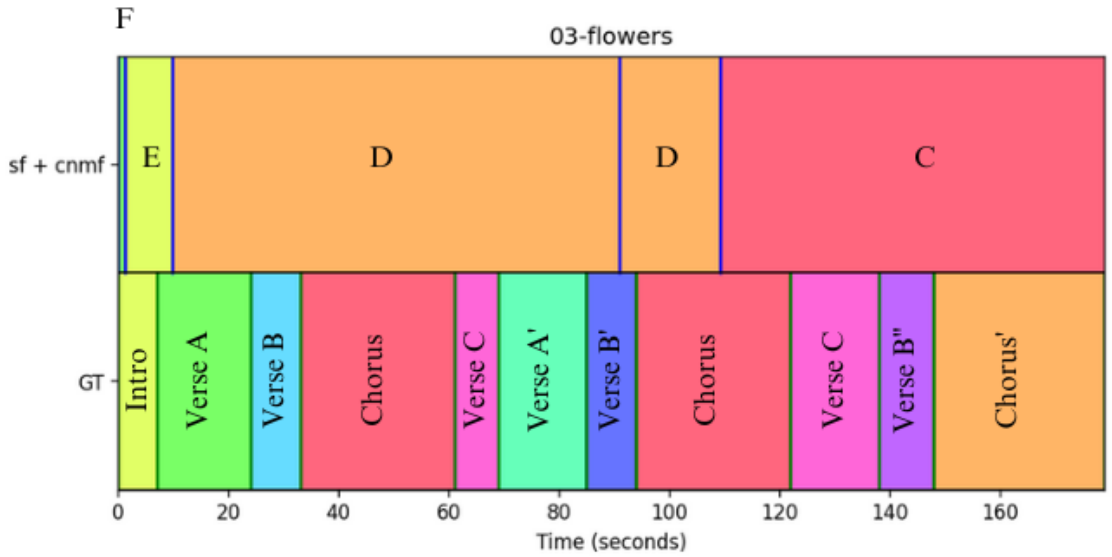


**Figura 12:** Comparação entre as segmentações feitas pelo algoritmo *Convex Non-Negative Matrix Factorization* [3], indicada por "sf+cnmf" e as anotações feitas pelo músico especialista, indicadas por "GT", no decorrer da duração da música Flowers em segundos, no experimento B. **Fonte**: Autora, 2023.

### 5.2.4 Música La Bebe

A taxa de acerto atingida na música La Bebe foi de 52,12% (PWF). O algoritmo CNMF [3] dividiu a música em quatorze partes, enquanto o músico especialista realizou a divisão dela em doze partes (Figura 13).

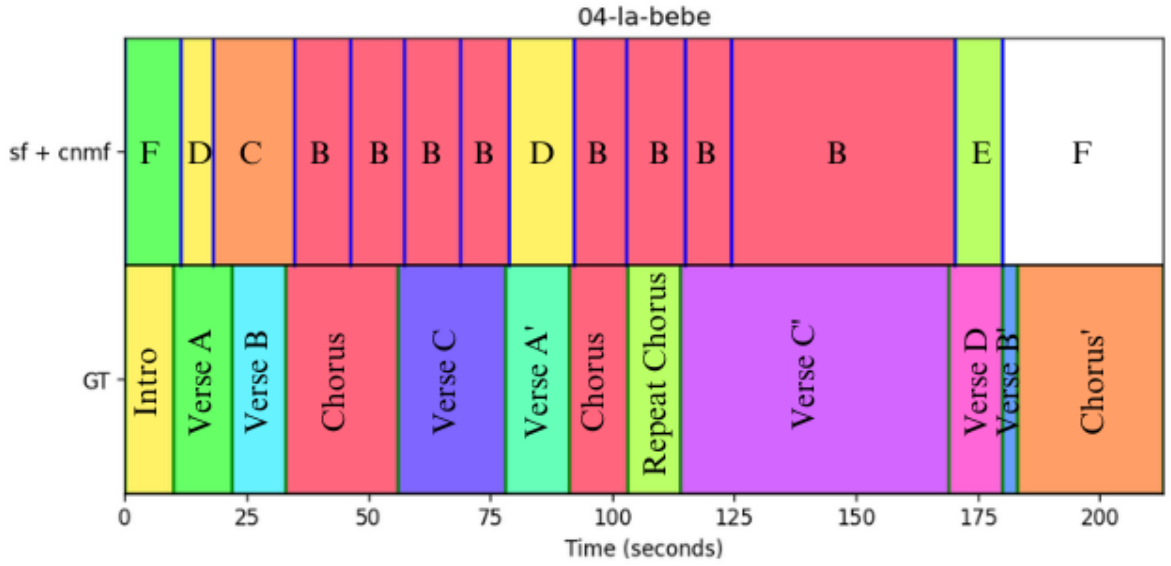



**Figura 13:** Comparação entre as segmentações feitas pelo algoritmo *Convex Non-Negative Matrix Factorization* [3], indicada por "sf+cnmf" e as anotações feitas pelo músico especialista, indicadas por "GT", no decorrer da duração da música La Bebe em segundos, no experimento B. **Fonte**: Autora, 2023.

### 5.2.5 Música unx100to

A taxa de acerto atingida na música unx100to foi de 47,37% (PWF). O algoritmo CNMF [3] dividiu a música em cinco partes, enquanto o músico especialista realizou a divisão dela em sete partes (Figura 14).

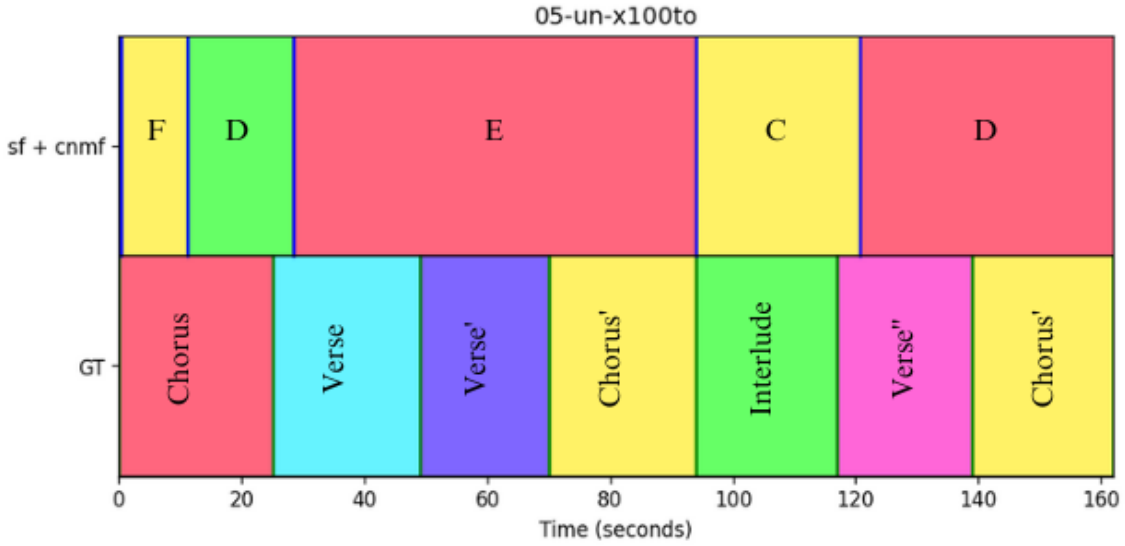



**Figura 14:** Comparação entre as segmentações feitas pelo algoritmo *Convex Non-Negative Matrix Factorization* [3], indicada por "sf+cnmf" e as anotações feitas pelo músico especialista, indicadas por "GT", no decorrer da duração da música unx100to em segundos, no experimento B. **Fonte**: Autora, 2023.

Tanto o algoritmo Fourier, quanto o CNMF, alcançaram os melhores resultados na música La Bebe. Como essa música foi discutida previamente, e as divisões das segmentações foram similares, a discussão será sobre a música Cupid, que apresentou o pior resultado no atual experimento. Quando essa música é comparada aos resultados obtidos no experimento A, as diferenças ocorrem entre 20 segundos e 1min30s de música. O algoritmo Fourier diferenciou a seção B da C, quando a vocalista principal sustentou a nota D (Ré maior) na frase "... *as a sign*", no fim do primeiro verso. Já o algoritmo CNMF realizou a demarcação de uma nova seção próxima aos 20 segundos, mas manteve a identificação do segmento como B, pois a progressão dos acordes Em7 (Mi menor com sétima), A7 (Lá maior com sétima), D (Ré maior), Bm7 (Si menor com sétima) e a dinâmica da música mantiveram-se basicamente iguais. Ambos os algoritmos só conseguiram identificar adequadamente os refrões na segunda parte da música.

## 5.3 EXPERIMENTOS A e B: FOURIER E CNMF COMPARADOS COM AS ANOTAÇÕES DO MÚSICO PROFISSIONAL

Em relação aos dois experimentos, o algoritmo *2D-Fourier Magnitude Coefficients* [8] alcançou resultados melhores em quatro músicas (Cupid, Ella Baila Sola, La Bebe e unx100to), enquanto o algoritmo *Convex Non-Negative Matrix Factorization* [3] apresentou um resultado melhor na música Flowers. Os dois experimentos foram executados levando em

consideração o mesmo algoritmo, *Structural Features* [10], que reconheceu os limites temporais das músicas, então a comparação foi feita em relação às segmentações das músicas utilizando PWF como principal métrica (Tabela 5).

| Música | Experimento Fourier (PWF) | Experimento CNMF (PWF) |
|---|---|---|
| Cupid | 47,51% | **36,92%** |
| Ella Baila Sola | 48,76% | 44,53% |
| Flowers | **36,52%** | 39,97% |
| La Bebe | **61,68%** | **52,12%** |
| unx100to | 55,95% | 47,37% |

**Tabela 5**: Comparação entre os resultados alcançados nos dois experimentos com o *framework* MSAF, levando em consideração como verdade base as anotações do músico profissional, com a métrica PWF. **Fonte**: Autora, 2023.

Em relação à música "Cupid", o melhor resultado atingido pelo MSAF [1] ocorreu no experimento que utilizou o algoritmo Fourier para a segmentação das músicas. Desconsiderando as diferenças de notações entre o algoritmo do MSAF [1] e o músico profissional, foram consideradas as marcações de tempo das seções para a análise qualitativa.

O algoritmo Fourier do MSAF [1] conseguiu acertar o que foi considerada a introdução da música pelo músico, entre 0 e 9 segundos, e acertou também a marcação de tempo de um trecho do verso A. Entretanto, o algoritmo englobou um trecho do verso A, o verso B completo e o primeiro refrão completo como se fosse apenas uma seção, o que não é intuitivo para os músicos. O músico profissional utilizou a notação (') para indicar que o trecho é similar a um que já aconteceu na música, mas que possui uma pequena diferença, seja na letra, ou até mesmo na dinâmica da música. Enquanto o algoritmo rotulou como uma seção só, o músico anotou o verso A' e verso B' entre aproximadamente 60 e 80 segundos da música.

Outro refrão acontece na música e o algoritmo conseguiu identificá-lo como uma seção nova, entretanto não acertou nem o começo nem o fim do carimbo de tempo que foi definido pelo músico. Em seguida, o músico profissional notou um interlúdio, uma passagem que conecta segmentos diferentes da música. O algoritmo também detectou uma nova seção, mas novamente houve um erro na marcação do tempo no começo e fim desta seção. O músico identificou um verso C, diferente dos demais, assim com o algoritmo MSAF [1], mas o

algoritmo errou novamente as marcações de tempo. Por fim, foi anotado o terceiro refrão pelo músico e o algoritmo rotulou corretamente o refrão, entretanto, o algoritmo anotou apenas dois refrões, enquanto o músico profissional notou três refrões.

Na música "Ella Baila Sola", o algoritmo Fourier implementado no MSAF [1] identificou um trecho rápido de 0,27s como uma seção E, no começo da música. O músico não anotou esta seção. Em seguida, o músico anotou a introdução e o algoritmo identificou uma nova seção B a partir dos 0,27s, que coincidiu com boa parte da marcação da introdução do músico. Em seguida, o músico anotou um verso e o algoritmo identificou uma nova seção A nesse mesmo trecho percebido pelo músico, mas esta seção do algoritmo terminou mais cedo do que o verso anotado pelo músico. Vale notar que o algoritmo repetiu a seção B, que foi considerada a introdução da música, na metade da seção classificada como verso pelo músico e até uma boa parte do que o músico avaliou como refrão da música.

Em seguida, em uma marcação de tempo próxima, tanto o algoritmo quanto o músico identificaram uma nova seção, anotada como "C" pelo algoritmo e anotada como "brass solo" pelo músico. Mas, para a percepção do músico, uma nova seção "verso(')" iniciou, enquanto a seção "C" ainda acontecia para o algoritmo. Por fim, o algoritmo englobou o refrão e o solo de guitarra anotados pelo músico como uma seção “D” e não conseguiu diferenciá-las, nem acertou o início exato dessa seção.

Na música "Flowers", o algoritmo Fourier do MSAF [1] apresentou o pior resultado do experimento, com apenas 36,52% na métrica PWF. O algoritmo criou uma breve seção E no começo da música, enquanto o músico classificou a introdução. Em seguida, o algoritmo anotou a seção C, que englobou boa parte da introdução percebida pelo músico. O algoritmo criou uma grande seção B que acabou englobando o que o músico classificou como verso A, verso B, refrão, pós-refrão, verso A' e verso B'. Em seguida, o algoritmo classificou a seção A, que coincidiu com parte do verso B' e do refrão do músico. Por fim, o algoritmo criou uma grande seção D que englobou parte do refrão, pós-refrão, verso B" e refrão'. Levando em consideração que as partes (seções) das músicas servem para facilitar a comunicação entre os músicos, podemos afirmar que o algoritmo não conseguiu separar as seções adequadamente e a sua usabilidade é quase nula para esta música.

Na música "La Bebe", o algoritmo Fourier do MSAF [1] apresentou o melhor resultado de todos os experimentos realizados, atingindo um resultado de 61,68%. Essa pontuação foi alta pois as marcações de tempo coincidiram com as anotações consideradas verdades fundamentais do músico profissional. Entretanto, vale ressaltar que, para este experimento, o algoritmo Fourier acabou utilizando bastante a anotação da seção "C", o que

deixa este tipo de segmentação utilizada contra intuitiva, para que os músicos utilizem esta nomenclatura para conseguirem se comunicar. Analisando visualmente o gráfico referente à música, as seções que mais se aproximaram na classificação foram as seções E e introdução, seção A com o verso A' e a seção D com a seção verso C" classificadas pelo algoritmo e músico, respectivamente.

Na música "Unx100to", o algoritmo Fourier do MSAF não conseguiu identificar nenhuma seção como o refrão da música, de acordo com o músico profissional. A única seção do algoritmo que iniciou na mesma marcação de tempo do músico foi o interlúdio, mas acabou um pouco depois do que o músico julgou ser o fim do interlúdio.

Em relação aos trabalhos relacionados, apesar de possuírem uma quantidade maior de músicas em seus conjuntos de dados e uma diversidade maior de estilos musicais, não houve uma discussão mais aprofundada sobre as diferenças entre as segmentações geradas pelos algoritmos e as segmentações feitas pelos músicos especialistas. Também não foram encontrados trabalhos que discutiram os resultados das músicas, uma por uma. Normalmente, são calculadas as taxas de acerto, com o auxílio de alguma ferramenta computacional como Matlab ou alguma linguagem de programação, mas não são evidenciadas as nuances entre os resultados dos algoritmos e as segmentações tidas como verdades fundamentais.

## 5.4 EXPERIMENTO C: MOISES, MÚSICO PROFISSIONAL E AUTORA

O Experimento C consistiu em realizar a comparação manual das classificações das segmentações de músicas feitas pelo aplicativo de música Moises e a comparação delas com as verdades fundamentais do músico profissional e autora. A fim de avaliar com mais facilidade, as segmentações do músico especialista e da autora foram convertidas para a mesma notação das segmentações do aplicativo Moises, como, por exemplo, seção "A", "B" e assim por diante. Além disso, nas seções que envolveram o refrão, foi feita uma simplificação, em que se o trecho da música for identificado como um pré-refrão, refrão ou pós-refrão, o trecho receberá apenas um rótulo de seção com uma letra designada para ele.

A comparação será feita em pares para facilitar a visualização, sempre começando pelo aplicativo comparado aos dados gerados pelo músico e depois aplicativo comparado aos dados gerados pela autora, para as músicas que tiverem uma estrutura mais complexa. Para simplificar a interpretação da comparação, na coluna do tempo das tabelas, o acordo total e uma diferença entre 1 e 4 segundos na marcação de tempo do segmento foram considerados

como acertos da segmentação, por parte do aplicativo Moises. Uma diferença maior ou igual a 5 segundos na marcação de tempo do segmento indicou um erro.

### 5.4.1 Música Cupid

Na música Cupid, o aplicativo Moises gerou 10 segmentações, enquanto a autora percebeu e anotou 12 segmentações (Figura 15). A única marcação temporal e rótulo que coincidiu entre os dois objetos de comparação foi feita na introdução da música. Em seguida, ocorreram 4 segmentações que se aproximaram em relação ao tempo, com diferenças entre 1 e 4 segundos. Por fim, um mesmo rótulo foi adotado, entretanto ele ocorreu em uma marcação temporal muito distante da avaliada pela autora, logo, é considerado um erro. Considerando o acordo total e com uma tolerância entre 1 e 4 segundos da marcação de tempo, apenas 4 das 12 segmentações coincidiram entre o aplicativo Moises e a autora (Tabela 6).

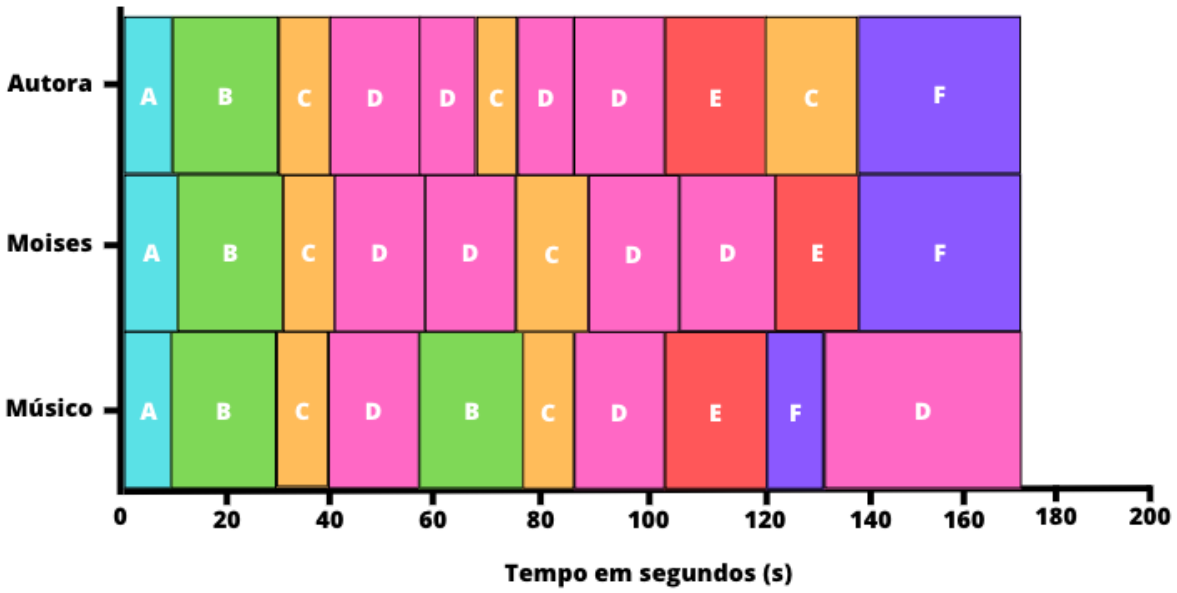



**Figura 15:** Comparação entre as segmentações feitas pelo Moises, as anotações feitas pela autora e pelo músico especialista, no decorrer da duração da música Cupid, em segundos, no experimento C. **Fonte**: Autora, 2023.

| Tempo Moises | Tempo Autora | Segmentação Moises | Segmentação Simplificada Autora |
|---|---|---|---|
| 00:00 | 00:00 | A | A |
| 00:10 | 00:09 | B | B |
| 00:26 | 00:25 | C | C |
| 00:42 | 00:39 | D | C |
| 00:58 | 00:57 | D | D |
| 01:14 | 01:05 | C | B |
| 01:30 | 01:13 | D | C |
| 01:46 | 01:27 | D | C |
| 02:02 | 01:45 | E | E |
| 02:18 | 02:01 | F | D |
| - | 02:17 | - | C |
| - | 02:52 | - | F |

**Tabela 6**: Comparação manual realizada entre a segmentação do aplicativo Moises e a da autora na música Cupid. **Fonte**: Autora, 2023.

A comparação entre o aplicativo Moises e o músico profissional atingiu melhores resultados do que a comparação realizada previamente entre o aplicativo e a autora, de acordo com os resultados apresentados na Tabela 7. Em relação à marcação de tempo, apenas a introdução apresentou um acerto total. No decorrer de toda a música, as marcações de tempo foram bem similares, não ultrapassando a tolerância entre 1 e 4 segundos. Em relação à percepção e marcação das segmentações, o Moises acertou 7 segmentos de 11 percebidos pelo músico, resultando em 77,77% (PWF).

| Tempo Moises | Tempo Profissional | Segmentação Moises | Segmentação Simplificada Profissional |
|---|---|---|---|
| 00:00 | 00:00 | A | A |
| 00:10 | 00:09 | B | B |
| 00:26 | 00:25 | C | C |
| 00:42 | 00:39 | D | D |
| 00:58 | 00:57 | D | B |
| 01:14 | 01:12 | C | C |

| | | | |
|---|---|---|---|
| 01:30 | 01:27 | D | D |
| 01:46 | 01:45 | D | E |
| 02:02 | 02:01 | E | F |
| 02:18 | 02:17 | F | D |
| - | 02:33 | - | D |

**Tabela 7**: Comparação manual realizada entre a segmentação do aplicativo Moises e a do músico profissional na música Cupid. **Fonte**: Autora, 2023.

Por fim, a comparação entre o músico e a autora apresentou muitas discordâncias sobre a classificação da segmentação da música, conforme apresentado na Tabela 8. Houve concordância total, tanto na marcação de tempo, quanto na classificação dos 3 segmentos, no começo da música. Após isso, houve concordância sobre novos segmentos aos 39 e 57 segundos da música, mas foram utilizados rótulos diferentes para classificar a música.

| **Tempo Autora** | **Tempo Profissional** | **Segmentação Simplificada Profissional** | **Segmentação Simplificada Autora** |
|---|---|---|---|
| 00:00 | 00:00 | A | A |
| 00:09 | 00:09 | B | B |
| 00:25 | 00:25 | C | C |
| 00:39 | 00:39 | D | C |
| 00:57 | 00:57 | B | D |
| 01:05 | 01:12 | C | B |
| 01:13 | 01:27 | D | C |
| 01:27 | 01:45 | E | C |
| 01:45 | 02:01 | F | E |
| 02:01 | 02:17 | D | D |
| 02:17 | 02:33 | D | C |
| 02:52 | - | - | F |

**Tabela 8**: Comparação manual realizada entre a segmentação da autora e a do músico profissional na música Cupid. **Fonte**: Autora, 2023.

### 5.4.2 Música Ella Baila Sola

Por se tratar de uma música estruturalmente mais simples, foi realizada a comparação do aplicativo Moises ao mesmo tempo com as anotações do músico profissional e da autora

do trabalho (Tabela 9). Em relação às marcações de tempo das segmentações, o aplicativo Moises coincidiu com apenas uma anotação do músico e da autora. Ao analisar este fator, o PWF foi considerado como 22,22% para esta música. Apesar de muitos rótulos do aplicativo Moises terem sido similares aos adotados pelos músicos, eles não foram identificados no tempo correto no decorrer da música. Já em relação à autora e o músico, 5 segmentos de 8 foram acordados (Figura 16).

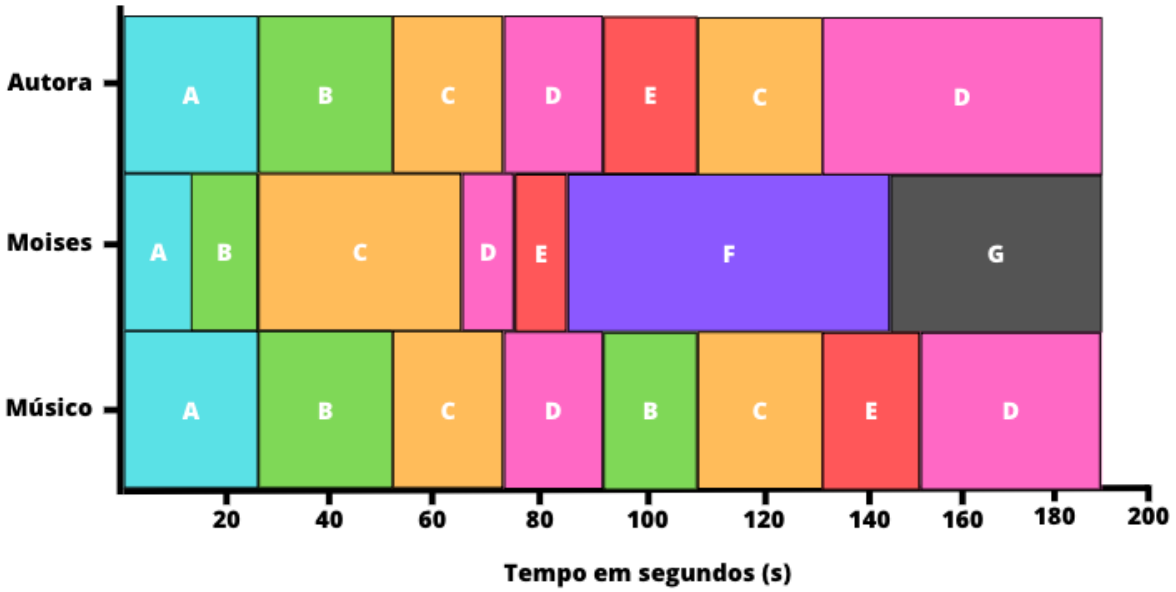



**Figura 16:** Comparação entre as segmentações feitas pelo Moises, as anotações feitas pela autora e pelo músico especialista, no decorrer da duração da música Ella Baila Sola, em segundos, no experimento C. **Fonte**: Autora, 2023.

| Tempo Moises | Tempo Autora | Tempo Profissional | Segmentação Moises | Segmentação Simplificada Autora | Segmentação Simplificada Profissional |
|---|---|---|---|---|---|
| 00:01 | 00:00 | 00:00 | A | A | A |
| 00:15 | 00:23 | 00:24 | B | B | B |
| 00:25 | 00:54 | 00:54 | C | C | C |
| 01:04 | 01:12 | 01:13 | D | D | D |
| 01:14 | 01:33 | 01:32 | E | E | B |
| 01:23 | 01:52 | 01:52 | F | C | C |
| 02:22 | 02:10 | 02:11 | G | D | E |

| | | | | | |
|---|---|---|---|---|---|
| - | - | 02:21 | - | - | D |

**Tabela 9**: Comparação manual realizada entre a segmentação da autora, a do músico profissional e a do aplicativo Moises na música Ella Baila Sola. **Fonte**: Autora, 2023.

### 5.4.3 Música Flowers

Na música Flowers, também foi realizada a comparação das segmentações do aplicativo Moises, do músico e da autora. Um erro que ocorre logo no começo da classificação do aplicativo pode comprometer quase totalmente a comparação feita com alguma verdade fundamental de um músico (Figura 17). O aplicativo Moises definiu que a introdução começou aos 2 segundos da música, o que acabou gerando um efeito cascata nas demais marcações de tempo, afetando o resultado final. Houve concordância com duas marcações de tempo entre o aplicativo e a autora, mas foram avaliados rótulos de segmentos distintos. A performance do aplicativo foi de 20% (PWF) em relação ao músico profissional. Já entre o músico e a autora, 4 segmentos foram definidos em acordo, com uma pequena variação no tempo da segunda marcação que ocorreu na música (Tabela 10).

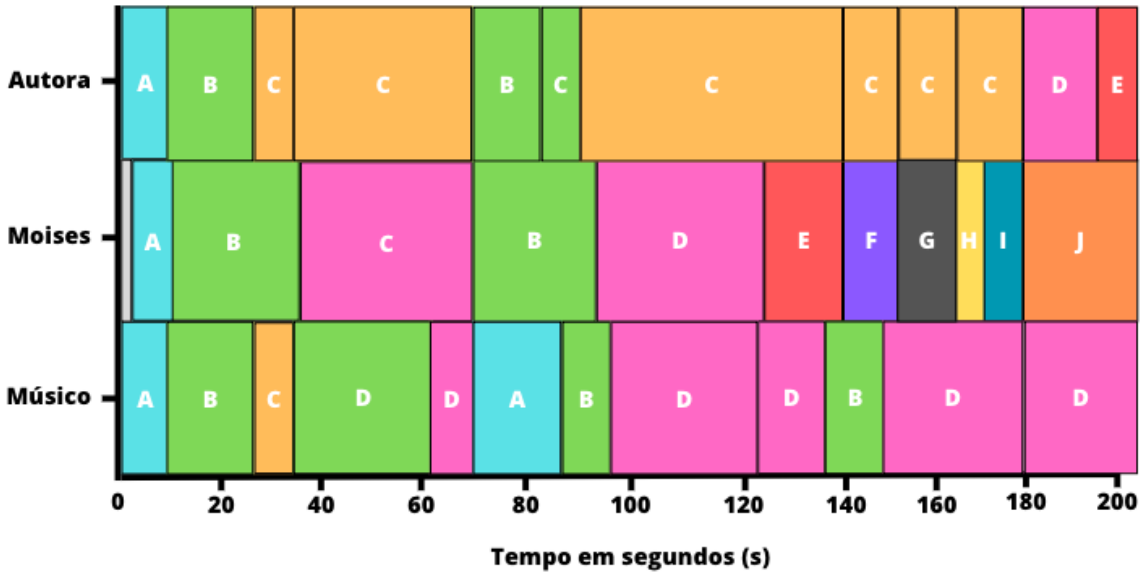



**Figura 17:** Comparação entre as segmentações feitas pelo Moises, as anotações feitas pela autora e pelo músico especialista, no decorrer da duração da música Flowers, em segundos, no experimento C. **Fonte**: Autora, 2023.

| Tempo Moises | Tempo Autora | Tempo Profissional | Segmentação Moises | Segmentação Simplificada Autora | Segmentação Simplificada Profissional |
|---|---|---|---|---|---|
| 00:02 | 00:00 | 00:00 | A | A | A |
| 00:09 | 00:08 | 00:07 | B | B | B |
| 00:35 | 00:24 | 00:24 | C | C | C |
| 01:10 | 00:33 | 00:33 | B | C | D |
| 01:34 | 01:10 | 01:01 | D | B | D |
| 02:03 | 01:26 | 01:09 | E | C | A |
| 02:19 | 01:33 | 01:25 | F | C | B |
| 02:27 | 02:19 | 01:34 | G | C | D |
| 02:44 | 02:27 | 02:02 | H | C | D |
| 02:50 | 02:52 | 02:18 | I | C | B |
| 03:00 | 03:00 | 02:28 | J | D | D |
| - | 03:16 | 02:59 | - | E | D |

**Tabela 10**: Comparação manual realizada entre a segmentação da autora, a do músico profissional e a do aplicativo Moises na música Flowers. **Fonte**: Autora, 2023.

### 5.4.4 Música La Bebe

Na música La Bebe, o algoritmo definiu a introdução da música apenas aos 3 segundos, assim como na música anterior. A hipótese está diretamente relacionada ao algoritmo que o aplicativo utiliza, que provavelmente só se baseia nas batidas das músicas para realizar esta segmentação, logo, se não há nenhum tipo de sinal no começo da música, o algoritmo trata como silêncio e não contabiliza o começo da marcação do tempo a partir de 0s (Figura 18). Trata-se de uma hipótese pois não foi divulgado abertamente pelo aplicativo qual é a técnica utilizada para a segmentação das músicas. Em relação ao aplicativo Moises e o músico profissional, 4 de 13 segmentos foram acertados 47,05% (PWF), levando em consideração principalmente a marcação do tempo. Na Tabela 11, houve um acordo entre a autora e o músico de 8 de 13 segmentações que foram anotados pelo músico.

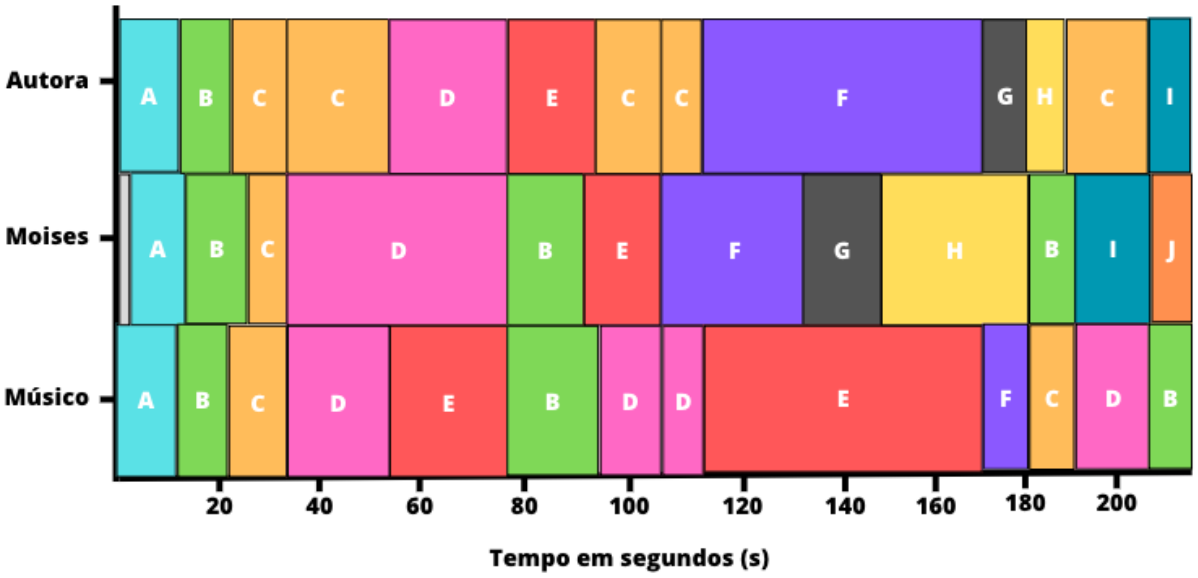



**Figura 18:** Comparação entre as segmentações feitas pelo Moises, as anotações feitas pela autora e pelo músico especialista, no decorrer da duração da música La Bebe, em segundos, no experimento C. **Fonte**: Autora, 2023.

| Tempo Moises | Tempo Autora | Tempo Profissional | Segmentação Moises | Segmentação Simplificada Autora | Segmentação Simplificada Profissional |
|---|---|---|---|---|---|
| 00:03 | 00:00 | 00:00 | A | A | A |
| 00:11 | 00:11 | 00:10 | B | B | B |
| 00:26 | 00:22 | 00:22 | C | C | C |
| 00:34 | 00:33 | 00:33 | D | C | D |
| 01:19 | 00:56 | 00:56 | B | D | E |
| 01:33 | 01:19 | 01:18 | E | E | B |
| 01:45 | 01:32 | 01:31 | F | C | D |
| 02:16 | 01:43 | 01:43 | G | C | D |
| 02:27 | 01:53 | 01:54 | H | F | E |
| 03:01 | 02:50 | 02:49 | B | G | F |
| 03:15 | 03:01 | 03:00 | I | H | C |
| 03:35 | 03:13 | 03:03 | J | C | D |
| - | 03:34 | 03:33 | - | I | B |
| - | 03:48 | - | - | J | - |

**Tabela 11**: Comparação manual realizada entre a segmentação da autora, a do músico profissional e a do aplicativo Moises na música La Bebe. **Fonte**: Autora, 2023.

### 5.4.5 Música unx100to

Na Tabela 12, foram trazidas as segmentações originais do músico e da autora, para demonstrar que mesmo utilizando notações originais diferentes, o objetivo de segmentar a música em partes iguais foi acordado em 100% entre os dois. Já em relação ao aplicativo Moises, considera-se que apenas a primeira e segunda segmentações foram acordadas com o músico e a autora, resultando em 44,44% (PWF) (Figura 19). O aplicativo Moises não conseguiu identificar o que os músicos identificaram como "Refrão" a partir de 01:10 da música, o que comprometeu o resultado final da classificação, levando em consideração a verdade fundamental do músico e da autora.

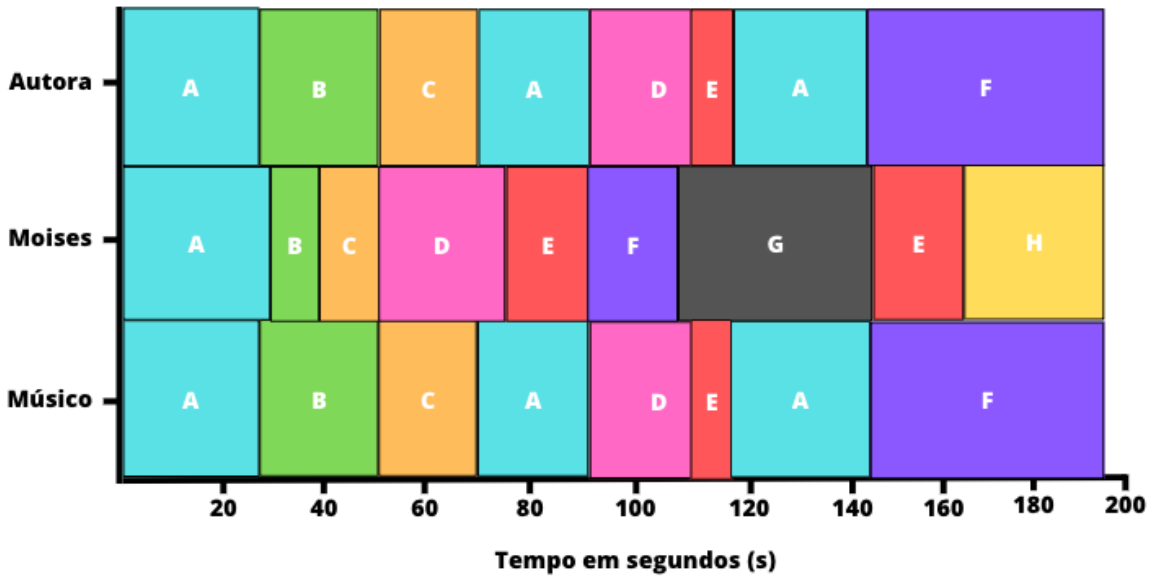



**Figura 19:** Comparação entre as segmentações feitas pelo Moises, as anotações feitas pela autora e pelo músico especialista, no decorrer da duração da música unx100to, em segundos, no experimento C. **Fonte**: Autora, 2023.

| Tempo Moises | Tempo Autora | Tempo Profissional | Segmentação Moises | Simplificada Autora | Simplificada Profissional | Original Autora | Original Profissional |
|---|---|---|---|---|---|---|---|
| 00:01 | 00:00 | 00:00 | A | A | A | Chorus | Chorus |
| 00:27 | 00:25 | 00:25 | B | B | B | Verse A | Verse |
| 00:38 | 00:49 | 00:49 | C | C | C | Verse B | Verse' |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 00:50 | 01:10 | 01:10 | D | A | A | Chorus | Chorus' |
| 01:16 | 01:35 | 01:34 | E | D | D | Interlude | Interludio |
| 01:35 | 01:58 | 01:57 | F | E | E | Verse C | Verse '' |
| 01:47 | 02:20 | 02:19 | G | A | A | Chorus | Chorus' |
| 02:24 | 02:42 | 02:42 | E | F | F | Solo | Solo |
| 02:43 | - | - | H | - | - | - | - |

**Tabela 12**: Comparação manual realizada entre a segmentação da autora, a do músico profissional e a do aplicativo Moises na música unx100to. **Fonte**: Autora, 2023.

Recapitulando os resultados obtidos no Experimento C, a música Cupid atingiu 77,77% (PWF) entre o músico e o aplicativo. A música Ella Baila Sola alcançou apenas 22,22% de acordo entre os dois. A música Flowers alcançou o pior resultado do experimento, com apenas 20% de acordo. A música La Bebe alcançou 47,05% e por fim a música unx100to conseguiu 44,44% (Tabela 13).

| **Música** | **Experimento Moises (PWF)** |
|---|---|
| Cupid | 77,77% |
| Ella Baila Sola | 22,22% |
| Flowers | 20,00% |
| La Bebe | 47,05% |
| unx100to | 44,44% |

**Tabela 13:** Comparação entre os resultados alcançados no experimento com o aplicativo Moises, levando em consideração como verdade fundamental as anotações do músico profissional, com a métrica PWF. **Fonte**: Autora, 2023.

Considerando os experimentos utilizando o MSAF e os experimentos do Moises, comparando a métrica PWF, que leva em consideração a precisão e o *recall*, o aplicativo Moises alcançou o melhor resultado na música intitulada Cupid, mas os demais resultados foram superados pelo algoritmo Fourier, executado no *framework* MSAF (Tabela 14). É importante frisar que não é possível afirmar que o MSAF apresentará um resultado melhor se eventualmente for integrado ao Moises, pois foi feita uma análise de uma quantidade pequena de músicas de estilo popular.

É necessário realizar testes e análises em músicas com estruturas mais complexas, para então afirmar com mais confiança que o MSAF traz resultados mais satisfatórios do que o Moises, em relação à classificação e segmentação das músicas. Outro ponto a destacar também é a praticidade do aplicativo Moises: no dia a dia, para os músicos, a usabilidade do aplicativo Moises é bem mais intuitiva e fácil do que o *framework* MSAF. Nem todos os músicos dominam Python para utilizar o *framework* e extrair os resultados do MSAF.

| **Música** | **Experimento MSAF Fourier (PWF)** | **Experimento Moises (PWF)** |
|---|---|---|
| Cupid | 47,51% | 77,77% |
| Ella Baila Sola | 48,76% | 22,22% |
| Flowers | 36,52% | 20,00% |
| La Bebe | 61,68% | 47,05% |
| Unx100to | 55,95% | 44,44% |

**Tabela 14**: Comparação entre os resultados alcançados no experimento com o MSAF, algoritmo Fourier, e com o aplicativo Moises, levando em consideração como verdade fundamental as anotações do músico profissional, com a métrica PWF. **Fonte**: Autora, 2023.

# 6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

O algoritmo disponível no *framework* MSAF, utilizado para classificar os limites das músicas nos 3 primeiros experimentos, gerou segmentações diferentes do músico especialista. Além disso, a notação utilizada pelo algoritmo do MSAF e pelo músico especialista divergiu, o que afetou também o resultado final dos experimentos realizados.

O aplicativo Moises apresentou resultados distintos do esperado mesmo que tenha ocorrido uma normalização e simplificação das notações das segmentações utilizadas para representar as avaliações do músico profissional e da autora. Apesar disso, o objetivo do trabalho de criar uma base de dados confiável e disponível para a comunidade científica e interessados na área de recuperação de informações musicais foi atingido.

Um complicador desta atividade foi o nível de discordância, ao realizar uma atividade subjetiva como a segmentação de músicas, pois, a título de exemplo, a autora do trabalho dividiu a canção Flowers em doze partes, enquanto o algoritmo do MSAF a dividiu em apenas seis partes, o músico especialista a segmentou em onze partes e o aplicativo Moises a dividiu em onze partes. Logo, fica evidente que a atividade proposta do experimento é complexa para normalizar e então mensurar.

A informática musical tem muito a crescer com o emprego da inteligência artificial, principalmente com a aprendizagem profunda de máquina, que pode ser amplamente aplicada nesta área de pesquisa. Como trabalho futuro, é possível desenvolver e treinar um modelo computacional capaz de segmentar as músicas, tanto na identificação dos limites temporais, quanto na classificação em si das partes das músicas.

É importante investigar com mais profundidade o emprego de outros tipos de algoritmos, que possam trazer uma taxa de acerto maior na identificação dos limites das músicas. Também é de grande relevância criar um sistema de padronização da notação utilizada para segmentar as músicas, tanto pelos músicos, quanto como o resultado dos algoritmos. Uma vez, com essa notação padronizada, envolver mais músicos no processo de avaliação e também calcular a discordância entre as segmentações criadas.

Adicionalmente, fazer os experimentos futuros levando em consideração um *threshold* de acordo com os compassos das músicas, para trazer flexibilidade nas marcações dos limites temporais que o algoritmo fará a predição, haja vista que há uma discordância natural entre músicos neste aspecto.

# 7 REFERÊNCIAS

[1] Nieto, O., Bello, J. P., *Systematic Exploration Of Computational Music Structure Research*. Proc. of the 17th International Society for Music Information Retrieval Conference (ISMIR). New York City, NY, USA, 2016.

[2] Mark Levy e Mark Sandler. *Structural Segmentation of Musical Audio by Constrained Clustering*. IEEE Transactions on Audio, Speech, and Language Processing, 16(2):318–326, Fevereiro de 2008.

[3] Oriol Nieto e Tristan Jehan. *Convex Non-Negative Matrix Factorization For Automatic Music Structure Identification*. In Proc. of the 38th IEEE International Conference on Acoustics Speech and Signal Processing, pages 236–240, Vancouver, Canada, 2013.

[4] Brian McFee e Daniel P. W. Ellis. *Analyzing Song Structure with Spectral Clustering*. In Proc. of the 15th International Society for Music Information Retrieval Conference, pages 405–410, Taipei, Taiwan, 2014.

[5] Matthew E. P. Davies, Norberto Degara e Mark D. Plumbley. *Evaluation Methods for Musical Audio Beat Tracking Algorithms*. Technical Report C4DM-TR-09-06, 8 de outubro de 2009.

[6] Jordan B. Smith, J. Ashley Burgoyne, Ichiro Fujinaga, David De Roure, and J. Stephen Downie. *Design and Creation of a Large-Scale Database of Structural Annotations*. In Proc. of the 12th International Society of Music Information Retrieval, pages 555–560, Miami, FL, USA, 2011.

[7] Eric J. Humphrey, Justin Salamon, Oriol Nieto, Jon Forsyth, Rachel M. Bittner e Juan P. Bello. *JAMS: A JSON Annotated Music Specification for Reproducible MIR Research*. In Proc. of the 15th International Society for Music Information Retrieval Conference, pages 591–596, Taipei, Taiwan, 2014.

[8] Oriol Nieto e Juan Pablo Bello. *Music Segment Similarity Using 2D-Fourier Magnitude Coefficients*. In Proc. of the 39th IEEE International Conference on Acoustics Speech and Signal Processing, pages 664–668, Florence, Italy, 2014.

[9] Colin Raffel, Brian Mcfee, Eric J. Humphrey, Justin Salamon, Oriol Nieto, Dawen Liang e Daniel P. W. Ellis. *mir eval: A Transparent Implementation of Common MIR Metrics*. In Proc. of the 15th International Society for Music Information Retrieval Conference, pages 367–372, Taipei, Taiwan, 2014.

[10] Joan Serra, Meinard Muller, Peter Grosche e Josep Lluis Arcos. *Unsupervised Music Structure Annotation by Time Series Structure Features and Segment Similarity*. IEEE Transactions on Multimedia, Special Issue on Music Data Mining, 16(5):1229 – 1240, 2014.

[11] Nieto, O., et al. (2020). *Audio-Based Music Structure Analysis: Current Trends, Open Challenges, and Applications*. Transactions of the International Society for Music Information Retrieval, 3(1), pp. 246–263, 2020.

[12] Lloyd, S. P. (1957). Least squares quantization in PCM. Technical Report RR-5497, Bell Lab, Setembro de 1957.

[13] McCallum, Matthew C. (2019). *Unsupervised Learning of Deep Features For Music Segmentation*. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Maio de 2019.