

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA POLITÉCNICA
DEPARTAMENTO DE ENGENHARIA ELETRÔNICA E DE
COMPUTAÇÃO

PROPOSTA DE PROJETO DE GRADUAÇÃO

Aluno: Lucas Lago de Carvalho
llagocarvalho@poli.ufrj.br

Orientador: Luiz Wagner Pereira Biscainho

1. TÍTULO

Processamento Digital de Áudio Aplicado à Detecção de *Samples* Musicais.

2. ÊNFASE

Processamento digital de áudio.

3. TEMA

O tema do trabalho é a detecção de trechos de uma determinada música (*samples*) em uma segunda música alvo. O trabalho consiste em implementar um sistema automatizado tal que, fornecidas como entrada uma música de referência e uma música suspeita, será capaz de determinar se existe alguma parte comum às duas.

4. DELIMITAÇÃO

O trabalho se restringe a analisar músicas gravadas em mono, ou seja, com apenas um canal, em formato digital de alta qualidade, como por exemplo, a uma taxa de 44,1 kHz (qualidade de CD). A base de dados utilizada consiste em 80 pares de músicas originais e músicas alvo. A duração do trecho pode variar de meio segundo a 20 segundos, sendo a média igual a 4,5 segundos.

Em geral, as músicas alvo são de hip-hop, principal gênero musical em que a técnica do *sampling* é empregada, e as músicas originais são em sua maioria funks dos anos 60 e 70, além de músicas pop dos anos 80. Na base de dados encontram-se casos em que o trecho original está presente na música alvo, mas com alteração de altura e/ou de andamento, e casos em que o trecho foi regravado em estúdio. Além disso, o excerto pode estar misturado com os mais diversos tipos de emissões sonoras,

como as de instrumentos gerados sinteticamente, de voz cantada e de outros trechos pré-gravados.

5. JUSTIFICATIVA

Desde os tempos pré-históricos, a música sempre esteve presente na vida humana. Durante sua evolução formaram-se períodos bem determinados com características próprias, mas certos aspectos se mantiveram presentes durante as transições entre as épocas. Tendo isso em mente, este trabalho poderá servir como ferramenta de análise das influências que uma geração musical sofre das gerações passadas, possibilitando novas percepções sobre a nossa cultura.

Há de se considerar também, entre as aplicações do algoritmo, a detecção de plágio. Não é de hoje que a cópia não autorizada de criações musicais é um problema. A própria técnica de *sampling* pode ser considerada plágio em algumas situações. O presente trabalho não seria capaz de definir a existência de uma irregularidade, mas poderia ser utilizado em uma primeira etapa para buscar, em uma grande coleção de músicas, suspeitos de infringir direitos autorais.

Ademais, o fato de não existirem na literatura muitos trabalhos que abordem o problema da detecção de *samples*, torna a busca por sua solução mais instigante e desafiadora.

6. OBJETIVO

A meta será implementar um sistema automatizado proposto em [1] para a detecção de *samples* musicais em diferentes gravações. Para isso, serão empregadas diversas técnicas conhecidas na literatura, como NMF, DTW e *Random Forest*. A robustez do método quanto à alteração de altura e à contração e expansão do tempo também será testada por meio de experimentos com músicas contendo tais variações.

7. METODOLOGIA

Um caminho natural ao se realizar uma detecção de um sinal presente em uma mistura é primeiro separá-lo para depois localizá-lo. Na literatura, o método mais empregado na tarefa da separação de sinais no tempo é a NMF (*Non-Negative Matrix Factorization*)[2], que busca fatorar uma matriz de espectrograma (representação tempo-frequencial de um sinal) [3] em uma matriz (\mathbf{W}) contendo em suas colunas os padrões espectrais de cada fonte (no contexto musical, um instrumento, uma nota,

um *sample* etc.) e uma matriz (\mathbf{H}) contendo em suas linhas as ativações desses padrões. Neste trabalho, adotou-se uma das diversas variações da NMF, chamada PFNMF (*Partially Fixed NMF*), na qual algumas bases espectrais são previamente definidas e mantidas fixas durante a fatoração. Assim, o algoritmo busca apenas por suas ativações temporais. As outras linhas de \mathbf{H} e respectivas colunas de \mathbf{W} são processadas normalmente. Esta escolha de algoritmo se deve ao fato de termos em mãos a música original, que, ao ser fatorada previamente por uma NMF, irá fornecer uma base espectral para a realização da PFNMF da música alvo.

Em um caso ideal, em que o *sample* não sofreu nenhuma modificação, uma correlação cruzada entre as matrizes de ativação da NMF da música original e da PFNMF da música alvo já seria o suficiente para encontrar os possíveis momentos em que o excerto é executado. Entretanto, uma primeira dificuldade surge quando o trecho é espalhado no tempo, já que o mesmo acontece com suas ativações. Além disso, variações de altura escalam o eixo frequencial da matriz \mathbf{W} de acordo com o fator aplicado na mudança de *pitch*. Isso exige que novas matrizes \mathbf{W} sejam geradas a partir do escalamento do eixo frequencial da matriz \mathbf{W} original de semitom em semitom, para então selecionar, dentre elas, a melhor candidata.

A DTW (*Dynamic Time Warping*)[4] aparece como uma excelente solução para os dois problemas, pois produz como resultado uma matriz de alinhamento, da qual é possível extrair informações sobre a semelhança entre as matrizes \mathbf{H} . Assim, em um primeiro momento, ela é utilizada para selecionar o semitom correto (os outros candidatos são descartados). Depois, extrai-se dela algumas *features* que serão passadas a um classificador *Random Forest*. Este último algoritmo se faz necessário, pois a probabilidade final do *sample* estar presente na música alvo não pode ser definida a partir de uma única característica da matriz de alinhamento, exigindo uma complexidade maior no momento da classificação.

A cada etapa implementada, testes utilizando a base dados e as anotações fornecidas pelo autor de [1] serão realizados para corroborar a escolha das técnicas. Isso também irá garantir o bom funcionamento do código e permitirá uma melhor compreensão do processo de detecção.

8. MATERIAIS

Irá se utilizar o Matlab como principal ferramenta de análise de dados, além do Excel para organização do projeto em geral.

9. CRONOGRAMA

Fase 1: Levantamento de diferentes técnicas existentes para detecção de *samples*.

Fase 2: Aprofundamento teórico a partir do método escolhido.

Fase 3: Implementação do algoritmo.

Fase 4: Análise de desempenho por meio de experimentos diversos.

Fase 5: Documentação.

Fevereiro	Março	Abril	Maio	Junho	Julho	Agosto
Fase 1						
	Fase 2					
		Fase 3				
					Fase 4	
				Fase 5		

Referências Bibliográficas

- [1] S. Gururani and A. Lerch, “Automatic sample detection in polyphonic music,” in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, (Suzhou), pp. 264–271, International Society for Music Information Retrieval (ISMIR), Janeiro 2017.
- [2] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, pp. 788–791, Outubro 1999.
- [3] P. S. R. Diniz, E. A. B. da Silva, and S. L. Netto in *Digital Signal Processing: System Analysis and Design*, Cambridge University Press, 2nd ed., 2010.
- [4] M. Muller, “Dynamic time warping,” in *Information Retrieval for Music and Motion*, pp. 69–84, Berlin: Springer, 2007.

Rio de Janeiro, 21 de março de 2018

Lucas Lago de Carvalho - Aluno

Luiz Wagner Pereira Biscainho - Orientador