

# Laboratório VISGRAF

Instituto de Matemática Pura e Aplicada

## **Visualizacao de Musica**

*Leandro Cruz*  
*Luiz Velho (supervisor)*

Technical Report    TR-17-05    Relatório Técnico

April - 2017 - Abril

The contents of this report are the sole responsibility of the authors.  
O conteúdo do presente relatório é de única responsabilidade dos autores.

# Visualização de Música

Leandro Cruz

14 de Fevereiro de 2017

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Abordagem Sinestésica para Visualização de Música</b>	<b>3</b>
2.1	Representações de Música . . . . .	3
2.2	O Método . . . . .	5
2.3	Operador de Visualização Musical . . . . .	6
<b>3</b>	<b>Linguagem Visual para Visualização de Música</b>	<b>10</b>
3.1	A Linguagem . . . . .	11
3.2	Um exemplo de elementos dessa Linguagem . . . . .	13
3.3	Extensões dessa Linguagem . . . . .	15
<b>4</b>	<b>Conclusão e Trabalho Futuros</b>	<b>16</b>
4.1	Composição de Música baseada em Imagens . . . . .	16
4.2	Criação de Automática de Vídeo Clipes . . . . .	18
	<b>Bibliografia</b>	<b>20</b>

# Capítulo 1

## Introdução

Ao longo desse relatório, apresentarei uma breve discussão sobre as pesquisas sobre Visualização de Música desenvolvidas no VISGRAF, ao longo do ano de 2016, por Leandro Cruz, Vitor Guerra e Juliano Kestenberg, sob a supervisão de Luiz Velho.

Desde o início do século passado, existem diversos trabalhos que criam representações visuais relacionadas com uma música. Em geral, essas representações estão associadas com as notas que estão sendo tocadas em um instante de tempo ou de acordo com as frequências do sinal sonoro gerado pela música. Nessa pesquisa, estamos desenvolvendo duas abordagens que relacionam representações visuais e música. A primeira diz respeito a uma linguagem visual que aprimora o processo de compreensão de certos fenômenos musicais. A segunda consiste na utilização da música para controlar parâmetros de métodos procedurais para processamento de imagens.

No Capítulo 2, apresentarei uma breve descrição sobre nossa abordagem sinestésica para visualização de música. No Capítulo 3, apresentarei alguns resultados e ideias para continuação dessa pesquisa, criando uma linguagem visual para visualização de música. Finalmente, na conclusão, apontarei para trabalhos futuros, em particular: (i) Composição de Música baseada em imagens e (ii) Criação de Automática de Vídeo Clipes.

## Capítulo 2

# Abordagem Sinestésica para Visualização de Música

Sinestesia significa perceber dois estímulos simultaneamente. Ao longo dessa pesquisa, nos focamos em criar estímulos visuais capazes de reforçar estímulos sonoros, destacando a ocorrência de certos elementos em uma música que esteja sendo tocada. Para tal, apresentamos um conjunto de operadores de imagens capazes de criar estímulos visuais associados a certos elementos presentes na música.

A sinestesia do uso de elementos musicais combinados com fenômenos de vídeo podem ser usados para aprimorar processos em educação musical, tal como, alfabetização musical, percepção da ocorrência de certos elementos (como timbres, acordes, etc), treinamento de ouvido, etc. Além disso, essa abordagem também tem um grande apelo artístico, pois pode ser usada como ferramenta por VJs (Vídeo Jockeys) para aprimorar a apreciação musical, ou para complementar apresentações musicais.

### 2.1 Representações de Música

Podemos agrupar as representações musicais basicamente em duas categorias: uma que representa explicitamente o sinal da música, ou seja, o comprimento da onda por instante de tempo; e outra que apresenta uma descrição explícita sobre as notas

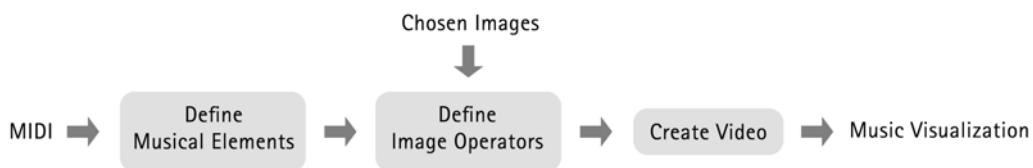


Figura 2.1: Pipeline da nossa abordagem para visualização musical.

que estão sendo tocadas por cada timbre a cada instante de tempo. Um arquivo de uma canção no formato MP3 se enquadra no primeiro modelo (salvo questões de codificação e compressão) enquanto um arquivo MIDI se encaixa no segundo.

Nas pesquisas que estamos desenvolvendo utilizamos uma reorientação descritiva das músicas (através dos arquivos MIDI, eventualmente adicionados de anotações referentes aos nossos procedimentos). Dessa forma, conhecemos a priori as notas tocadas em uma música, qual é seu respectivo timbre, em qual instante de tempo essa nota começou a ser tocada e em qual ela parou, e qual foi a velocidade em que a nota foi tocada.

Com as informações presentes nos arquivos MIDI, podemos definir os acordes como agrupamentos de notas que começaram a ser tocadas e terminaram nos mesmos respectivos tics. Sendo assim, nessa pesquisa, nos focaremos nesses elementos musicais: timbres, acordes, notas, e a velocidade em que a nota é tocada.

Existem outros elementos musicais, além dos identificados em um arquivo MIDI, que podem ser extraídos de uma música representada em um arquivo no formato de onda. Algumas dessas informações são obtidas utilizando técnicas de processamento de sinais, bastante conhecidas na literatura. Por exemplo, Lewiner et. al. [1] apresenta uma técnica de controle de objetos 3D a partir de uma análise espectral de uma música dada.

Apesar das inúmeras possibilidades provenientes de uma análise mais sofisticada de uma música, no momento, não abordamos técnicas de análises mais sofisticadas. Contudo, para trabalhos futuros com representações visuais mais complexas precisaremos identificar elementos musicais mais complexos: como ritmo, melodia, dinâmica e harmonia. Para tal, precisaremos de análises mais sofisticadas.

## 2.2 O Método

Esse método consiste em, dada uma música, criar cada frame de um vídeo aplicando operadores de imagens (técnicas clássicas ou específicas de processamento de imagens) no frame anterior, de acordo com a ocorrência de certos elementos musicais, em um conjunto de imagens pre-selecionadas. Nem todos os operadores irão produzir um efeito sinestésico desejado. A escolha das imagens e de tais operadores é feita combinando valores estéticos com a intenção de quais elementos musicais deseja-se destacar. Mas além disso, apresentamos um conjunto de princípios que guiam a forma de relacionar eventos na música com a respectiva representação visual.

A Figura 2.1 ilustra o pipeline da nossa abordagem. Em particular, para cada

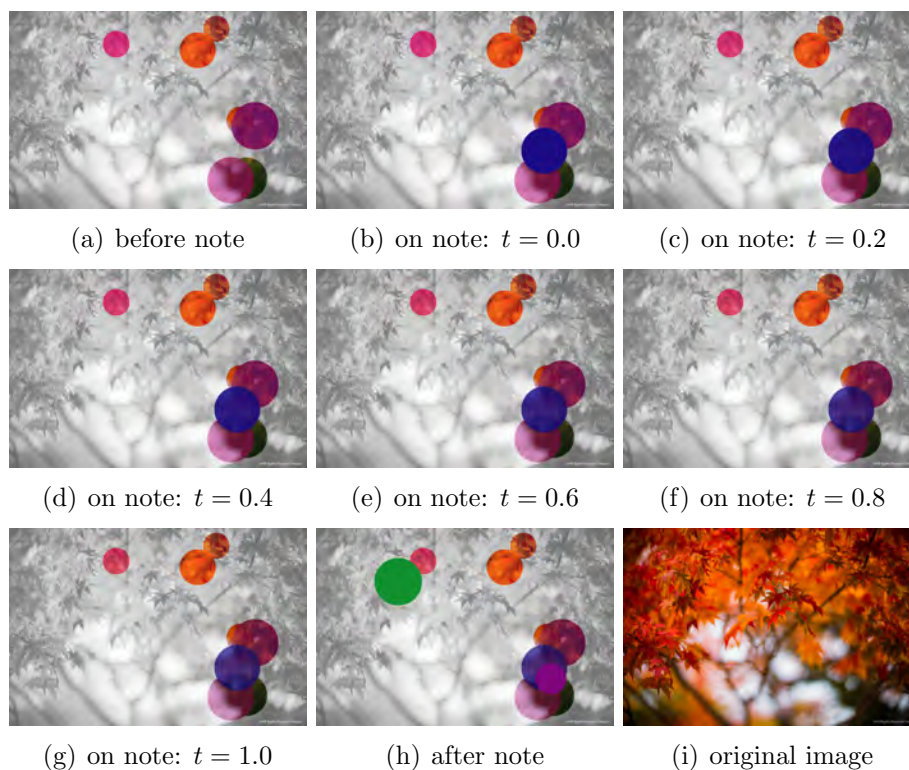


Figura 2.2: It illustrates the color variation derived from LCO: (b) has a blue circle (note color) that gradually changes to the color of the original image during the time the note is being played (from  $t = 0$  to 1.0).

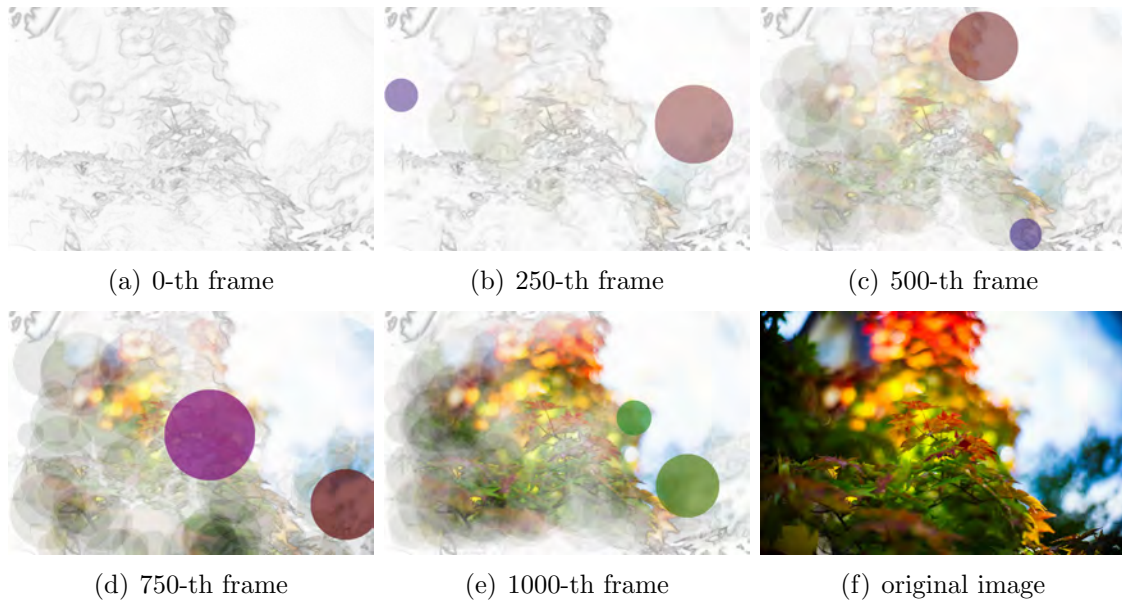


Figura 2.3: Exemplos de frames de uma visualização musical criada usando um operador do tipo RIO

frame gerado no vídeo, é necessário identificar o respectivo elemento musical que está sendo tocado no respectivo momento, e produzir um elemento gráfico correspondente. Cada elemento musical é usado para controlar um determinado efeito sobre as imagens.

## 2.3 Operador de Visualização Musical

Um operador de imagem é uma função que recebe uma imagem (ou várias) e retorna uma imagem. No nosso caso, esse operador é parametrizado pelos elementos musicais identificados na música. Tais operadores devem respeitar os seguintes princípios:

- Estabelecer uma associação entre cores das imagens e notas da música
- Estabelecer uma relação entre a quantidade de pixels que esteja sendo processada, a cada novo frame, com a velocidade das notas que estão sendo tocadas



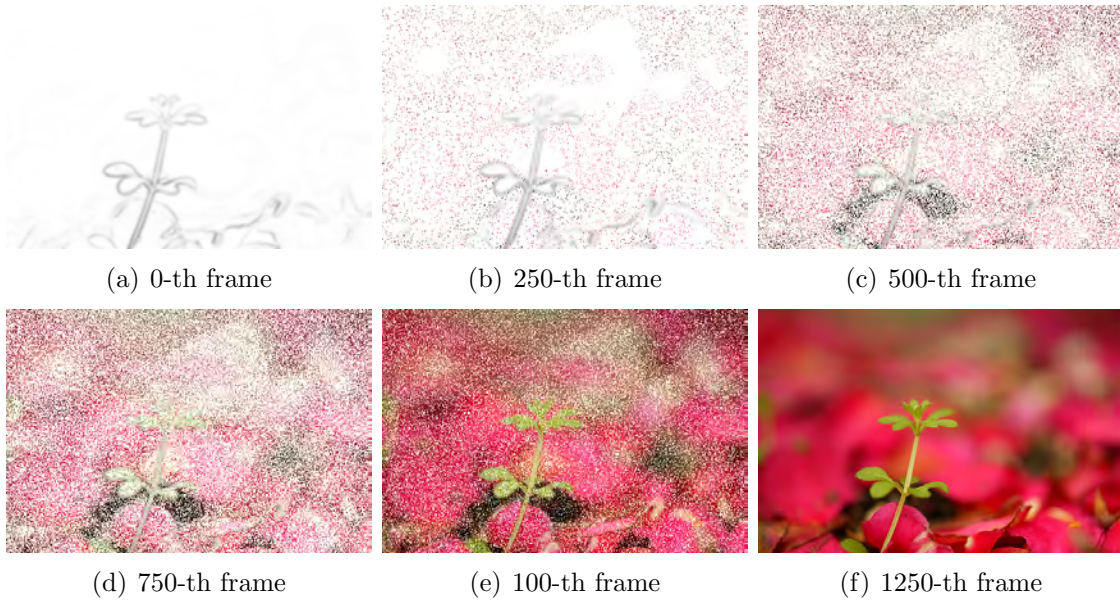


Figura 2.4: Exemplos de frames de uma visualização musical criada usando um operador do tipo PSO

- Dividir o frame em regiões distintas, cada uma associada a um determinado timbre

Esses princípios foram criados para garantir a sinestesia dos operadores. Além disso, são bastante gerais, e portanto podem ser facilmente extensíveis. Chamamos os operadores que respeitam tais princípios de VMO (Visual Music Operator, ou Operadores de Visualização de Música). Naturalmente, também podemos estender esses princípios para abarcar outros elementos musicais, ou mesmo, criar relações distintas entre os citados elementos musicais com efeitos visuais.

Nesse trabalho, buscamos apresentar um framework conceitual para o processo de visualização de uma música, através do uso de VMO, criados a partir de um conjunto de princípios sinestésicos coerentes. Validamos essa abordagem através da criação de quatro classes de VMOs seguindo tais princípios:

1. Luminance-Chrominance Operators (LCO)
2. Residual Info Operators (RIO)

### 3. Pixel Selection Operators (PSO)

### 4. Image Combination Operators (ICO)

Um trabalho futuro natural para essa pesquisa é estender esse conjunto de princípios para ser capaz de lidar com elementos musicais mais sofisticados, como ritmo, melodia, dinâmica e harmonia. Para tal, precisaremos melhorar o processo de análise de som, e definir novos princípios sinestésicos. Mas vale destacar que o pipeline de criação de visualização musical apresentado é perfeitamente adequado para essas futuras extensões. No Capítulo 3 comentaremos sobre direções de pesquisas que permitam estender essa abordagem proposta para a criação de visualizações musicais.

Também vale destacar que apesar do valor estético e semântico das imagens naturais, não encontramos nenhum outro trabalho que as tenham explorado satisfatoriamente na visualização de músicas. Uma linha muito interessante para trabalhos futuros dessa pesquisa é identificar o conteúdo semântico das imagens automaticamente ao longo da criação das visualizações das músicas. No Capítulo ?? apresentaremos uma possível abordagem para aproveitar o conteúdo semântico das imagens para o propósito de visualização musical.



(a) original background



(b) base background



(c) original foreground



(d) base foreground



(e)



(f)



(g)



(h)



(i)



(j)



(k)



(l)

Figura 2.5: Exemplos de frames de uma visualização musical criada usando um operador do tipo ICO

## Capítulo 3

# Linguagem Visual para Visualização de Música

Uma linguagem é caracterizada por um conjunto de regras sintáticas e seus efeitos semânticos. Ou seja, a sintaxe é um conjunto de regras sobre como usar os elementos da linguagem (as palavras). E a semântica é o significado que tiramos do uso de um conjunto de palavras postos consecutivamente (as frases) respeitando as regras sintáticas.

Neste capítulo apresentaremos uma proposta de pesquisa que visa definir um conjunto de regras sintáticas de uma linguagem que caracterizam como usar certas palavras visuais, criadas a partir de elementos de uma música, para representar visualmente certos fenômenos musicais (semântica). Assim como no trabalho apresentado no Capítulo 2, podemos começar a pensar sobre essa linguagem a partir de uma análise feita com elementos musicais simples, como timbres, acordes, as notas e a velocidade na qual uma nota é tocada. Entre os fenômenos musicais que podemos representar com essa linguagem estão, percepção musical (timbres e acordes), melodia, o ritmo e a harmonia.

Este tipo de linguagem pode ser amplamente utilizada tanto com propósitos pedagógicos (aprendizagem musical e como percepção de elementos), bem como com propósitos artísticos. Vale destacar, que a criação dessa linguagem visual é uma

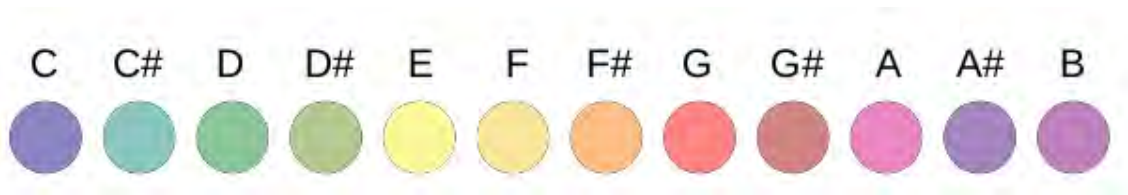


Figura 3.1: Paleta de cores de Louis Bertrand Castel.

extensão do processo sinestésico previamente apresentado. Pretendemos nos aprofundar na formalização das relações entre elementos musicais e elementos visuais, bem como, na construção de uma ferramenta que permita criar uma visualização musical com um amplo conjunto funcionalidades para tal fim.

### 3.1 A Linguagem

Para definirmos uma linguagem visual para visualização de música, precisamos inicialmente definir o que é uma linguagem visual.

Existe uma literatura ampla sobre criação de linguagens visuais para diversos propósitos. Em geral, essas linguagens consistem em um encadeamento lógico de comandos feitos em uma ferramenta interativa e visual (por exemplo, conectando caixinhas com significados específicos, através de regras específicas, definindo assim, um fluxo de processamento). Ou seja, esse tipo de linguagem é como uma linguagem de programação, onde as regras sintáticas são implementadas através de ferramentas visuais.

No nosso contexto, a linguagem visual segue uma abordagem diferente. Queremos criar uma forma de comunicação (sintaxe) que relacione elementos de duas naturezas distintas (visual e musical) de uma coerente e coerente (semântica). Para tal, precisamos entender melhor questões como:

1. Definir mais precisamente o que é sintaxe e semântica (de acordo com alguma fonte linguística)
2. caracterizar nossa sintaxe

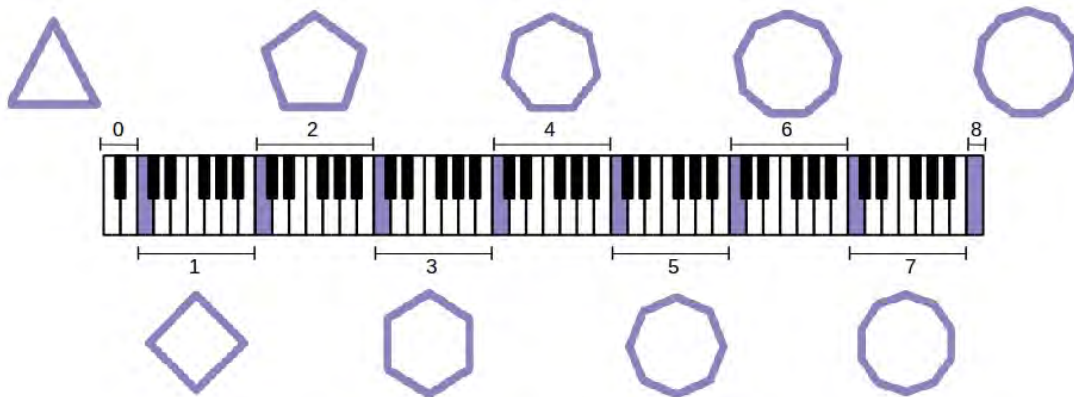


Figura 3.2: Exemplo de relação entre quantidade de lados do polígono e a oitava da respectiva nota

- (a) definir em termos gerais os tipos de representações visuais que podem ser gerados a partir dos elementos de uma música
  - (b) definir palavras e frases visuais
  - (c) definir representações que são sinônimos (com a mesma semântica)
3. justificar essa caracterização (usando questões como sinestesia, didática, e possivelmente estética)
  4. determinar quais são as implicações semânticas obtidas a partir da nossa linguagem (fenômenos musicais).
  5. fazer uma breve avaliação da eficácia semântica de nossa linguagem

Após a definição conceitual da linguagem, precisamos gerar uma série de experimentos que abordem elementos distintos para verificar o efeito sinestésico do uso de algumas representações visuais. Ou seja, é essencial fazer uma avaliação da eficácia semântica de nossa linguagem.

Vale destacar que uma abordagem factível para essa pesquisa é abordar essa caracterização usando apenas elementos musicais simples (como timbres e acordes) e mais adiante estender a abordagem para ser capaz de lidar com peças musicais



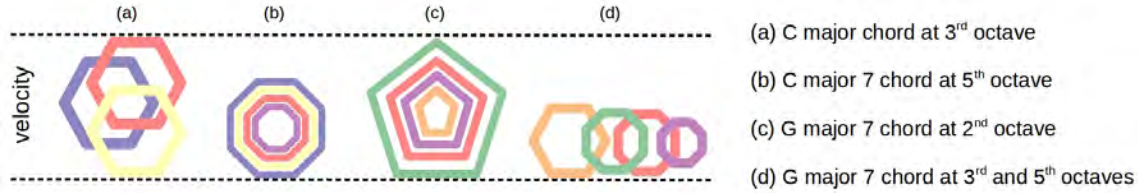


Figura 3.3: Exemplo de relação entre acordes e composição de polígonos

complexas, como uma ópera, por exemplo. Contudo, o caso teste já será suficiente para explorar a conceitualização linguística e o potencial dessa linguagem visual para a visualização de música.

## 3.2 Um exemplo de elementos dessa Linguagem

Os primeiros experimentos feitos nessa direção estabeleceram um conjunto de relações entre elementos visuais e elementos musicais (sem uma caracterização linguística). Nessa seção apresentaremos algumas dessas relações e apontaremos direções para aprimorá-las.

De um modo geral, esse exemplo de visualização de música é baseada na utilização de elementos visuais sintéticos (como por exemplo polígonos ou círculos) que possuem certas propriedades (cores, posição, tamanho, borda, etc) e certos comportamentos (transladar, escalar, borrar, etc). Cada forma geométrica será associada a uma nota que será tocada na música.

A primeira relação que utilizamos é uma relação amplamente conhecida em visualização de música: a relação de cores e notas, proposta por Louis Bertrand Castel em 1725 . Ele argumenta a existência de um efeito sinestésico entre os elementos dessa paleta. Adotamos essa paleta como um padrão nessa pesquisa, haja visto que tem sido utilizada em diversos outros trabalhos nessa linha [2].

Uma segunda relação é determinar a quantidade de lados desse polígono de acordo com a oitava da respectiva nota. A Figura 3.2 ilustra essa relação. Estabelecemos, também, que o tamanho do polígono é determinado em função da velocidade da



Figura 3.4: Exemplo de relação entre acordes e composição de polígonos

respectiva nota.

Além disso, precisamos definir a posição de tais formas. Partimos de um princípio semelhante ao que foi apresentado no Capítulo 2, de particionar o frame de acordo com o timbre: ou seja, exibir as representações visuais associadas a um timbre em uma parte específica do frame. Testamos posicionar as formas aleatoriamente dentro dessas regiões, bem como, determinar previamente tal posição. A adequação de cada escolha variou de caso para caso. A Figura 3.4 mostra um exemplo de uma música que está sendo tocada com quatro timbres. Nesse caso, o frame foi dividido em quatro partes, e as respectivas notas são mostradas no respectivo quadrante.

Essas regras sintáticas descritas para as representações visuais de uma nota são extremamente simples e cuja semântica visa apenas descrever o tipo da nota. Para definir a uma regra sintática de uma nota podemos aproveitar outras propriedades visuais de um polígono como desenhar sua curva de bordo, preencher ou não o interior, combinar interior com uma cor e borda com outra, ter um certo grau de transparência, etc.

Já um acorde, que é uma combinação de notas, é representado pela combinação das representações visuais das respectivas notas. A Figura 3.3 ilustra algumas formas de representar um acorde: aninhar as formas geométricas, ou colocá-las próximas com um alinhamento aleatório.



### 3.3 Extensões dessa Linguagem

Na seção anterior apresentamos algumas relações bastante básicas entre elementos musicais e visuais. Precisamos destacar que essa linguagem baseada apenas em timbres, acordes e notas é uma versão inicial dessa linguagem visual para representação de música, e que existem alguns elementos da música que não são bem representados, por exemplo, percussão, efeitos eletrônicos (distorção de guitarra, sons sintetizados) e voz. Estamos primeiramente avaliando o potencial dessa versão mais simples da linguagem, e em um futuro próximo pretendemos estendê-la para incluir tais elementos. Outra extensão necessária, será representar agrupamentos de instrumentos para, por exemplo, conseguir fazer uma representação mais compacta, e portanto mais compreensível, de uma música tocada por uma orquestra grande.

Também vale destacar que podemos criar diversas regras sintáticas sobre os elementos timbres, acordes e notas, e estabelecer uma regra semântica para cada uma delas. Existirão regras sinônimas, ou seja, caracterizações distintas associadas ao mesmo significado musical. Parte dessa pesquisa envolve em caracterizar adequadamente o uso de representações sinônimas.

Um grande desafio dessa pesquisa é apresentarmos uma caracterização que seja genérica o suficiente que nos permita criar diversas regras para criação de representações visuais associadas a elementos musicais, permitindo a existência de representações visuais distintas associadas ao mesmo fenômeno musical, mas que seja bem definida o suficiente, de modo que a interpretação semântica da relação esteja clara.

# Capítulo 4

## Conclusão e Trabalho Futuros

Nesse relatório, foi apresentado alguns resultados obtidos na pesquisa sobre visualização musical desenvolvida no VISGRAF, ao longo de 2016, por Leandro Cruz, em colaboração com Vitor Guerra e Juliano Kestenberg, sob a supervisão de Luiz Velho.

Um primeiro resultado dessa pesquisa foi o trabalho sobre a sinestesia na criação de visualização para música (A Synesthetic Approach for Music Visualization, artigo submetido e aguardando resposta). Além desse trabalho, também avançamos em outras linhas, com alguns resultados iniciais, e ideias para trabalhos futuros. O principal desdobramento foi o apresentado sobre a linguagem visual para visualização de música. Nesse capítulo, apresentaremos outras duas ideias, que embora ainda não tenham sido iniciadas, se encaixam organicamente com a linha que tem sido abordada.

### 4.1 Composição de Música baseada em Imagens

Nessa seção apresentarei uma ideia de trabalho futuro que ainda não foi nem iniciada. Essa ideia consiste basicamente em criar estruturas de controle para o processo de composição automática a partir de um conjunto de imagens. Mais especificamente, podemos partir de uma ou várias imagens, e definirmos alguns descritores específicos que sejam usados para controlar uma máquina de composição algorítmica de uma

música.

Entre as estruturas de controle, podemos citar a mais simples, como um histograma de cores das imagens, de acordo com a paleta de Castel (apresentada no Capítulo 3). Mais especificamente, dada uma coleção de imagens, para cada pixel, determinando qual é a cor da paleta de notas mais próxima. Em seguida, criamos um histograma associado às cores de paleta, determinando para cada cor, a respectiva quantidade de pixels associados. Finalmente, durante o processo de composição, precisaremos garantir que a música gerada tenha uma distribuição de notas equivalente ao histograma criado.

Naturalmente, essa primeira estrutura é muito simples, e determinar apenas a quantidade de cada notas de cada tipo não propicia um controle sofisticado. Uma possibilidade é além da distribuição das notas, também controlarmos a velocidade em que as notas são tocadas. Esse controle pode ser feito a partir de uma análise do tamanho das componentes conexas nas imagens associadas a respectiva nota.

Ou seja, após a associação de pixel com a respectiva nota, identificamos as componentes conexas associadas a cada nota, calculamos seu tamanho (quantidade de pixels) e na hora da composição adicionamos essa restrição. Esse controle já é mais restritivo, e pode permitir a criação de músicas bem diferentes, a medida que se muda a coleção de imagens.

Naturalmente, é necessário testar se esse tipo de controle gerará músicas coerentes e agradáveis. Vale destacar que esse controle é probabilístico, ou seja, adicionaremos uma componente na síntese que tende a favorecer a composição respeitando a informação proveniente da estrutura de controle. Dessa forma, uma estrutura de controle muito incompatível com uma determinada máquina de composição pode ter pouco efeito de controle. Uma possibilidade é que se arbitrarmos uma máquina, dada uma coleção de imagens, buscarmos uma máquina que seja razoavelmente semelhante (ou seja, que a probabilidade de gerar um resultado de acordo com a estrutura de controle é alta).

Não estamos enfatizando o processo de composição. Embora, ao longo dessa pesquisa, é possível que surjam ideias interessantes também relacionadas a criação das máquinas de composição (por exemplo, criar máquinas mais adequadas às estruturas

de controle propostas).

O objetivo central dessa proposta é criar uma experiência sinestésica, onde os participantes podem perceber mudanças na música que está sendo criada/tocada de acordo com a manipulação do conjunto de imagens que controla a máquina de composição. Dessa forma, uma possibilidade bastante interessante é criar uma instalação onde os usuários podem interagir de maneira natural com tais imagens (adicionar ou remover imagens, escalar imagens, ou fazer algum tipo de manipulação coerente com os propósitos desse trabalho) e perceber as mudanças geradas na música que está sendo gerada. Contudo, versões iniciais podem ser feitas sem tamanho aparato (apenas criando uma lista das imagens usadas, possíveis transformações, e com isso gerando as estruturas de controle).

Essa ideia nasceu relacionando ferramentas de computação gráfica, semelhantes às que foram usadas para criar visualizações para música, com a geração automática de músicas. Esse segundo ponto, tem sido o tema central da pesquisa do Vitor Guerra, ao longo do seu Pós-doutorado no VISGRAF. Dado que o processo de síntese já está sendo abordado pelo Vitor, parece-me que não há grandes obstáculos para o desenvolvimento dos passos iniciais desse trabalho (e verificação se de fato gerará resultados interessantes).

## 4.2 Criação de Automática de Vídeo Clipes

Recentemente a área de *Machine Learning* apresentou avanços significativos para reconhecimento de elementos em uma imagem, ou seja, atribuir semântica a uma imagem. Já há trabalhos que identificam muito eficientemente elementos em uma imagem. Uma tarefa um pouco mais complicada é estabelecer sentenças sobre essa imagem. Apesar de ser uma tarefa mais difícil, recentemente, também já surgiram alguns trabalhos apresentando bons resultados nessa direção.

Dessa forma, um possível trabalho futuro para criação de representações visuais para música é criar um vídeo clipe, mostrando uma sequência de imagens de acordo com os versos de uma música que esteja sendo tocada. Para tal, representaremos uma música como o arquivo de áudio, mais sua letra, sincronizando os versos com

o momento em ele é cantado (como uma legenda em um vídeo). Para cada verso, procuraremos uma imagem relacionada. Isso pode ser feito buscando por imagens relacionadas a sentenças semelhantes ao respectivo verso.

Apesar dos avanços de *Machine Learning*, essa pesquisa é bem mais desafiadora ao que foi descrito na seção anterior, dado que é necessário obter uma base de dados adequada, e conseguir fazer um casamento entre sentenças, o que pode ser não trivial. Contudo, os constantes avanços dessa área indicam que essa é uma possibilidade viável.

Além desse processo de escolha, podemos criar algumas ferramentas extras para valorizar o efeito visual do clipe. Entre essas ferramentas, podemos citar translações ou escalamento da imagem escolhida ao longo do frame. Além disso, podemos adicionar outros efeitos visuais semelhantes ao que foi apresentado no Capítulo 2.

# Bibliografia

- [1] T. Lewiner, C. Marques, J. Paixão, S. Botton, A. Cabral, R. Nascimento, V. Mello, A. Peixoto, D. Martinez, and T. Vieira, “Stereo music visualization through manifold harmonics,” *The Visual Computer*, vol. 27, no. 10, pp. 905–916, 2011.
- [2] K. Brougher, J. Strick, A. Wiseman, and J. Zilczer, *Visual Music: Synaesthesia in Art and Music Since 1900*. Thames and Hudson, 2005.