

Miguel
Cruz

Sistema de geração
de imagens de acordo
com texto e som para
álbuns de música

Faculdade de Ciências e Tecnologia
da Universidade de Coimbra

Orientação:
Paul Hardman
João M. Cunha
Tiago Cruz

Julho 2019



Os
olhos
também
ouvem



Miguel Machado Cruz

migmacruz@gmail.com

Os olhos também ouvem

Sistema de geração de imagens
de acordo com texto e som
para álbuns de música

Dissertação no âmbito do curso de
Mestrado em Design e Multimédia

Orientação:

Paul Hardman

João M. Cunha

Tiago Cruz

Júri:

João Bicker

Pedro Martins

Faculdade de Ciências e Tecnologia
da Universidade de Coimbra

Julho 2019

À minha mãe e ao meu pai. Tornaram-me quem sou e apoiaram-me incondicionalmente desde o meu primeiro dia. Devo-lhes o mundo!

À Rita. Ofereceu-me amor, carinho e paciência, tornando os momentos difíceis mais leves e os bons momentos melhores ainda. Transformou Coimbra na minha nova casa. É um desafio maravilhoso e inspirador partilhar uma vida com ela.

Ao Mário, à Inês e às luzes que deles nasceram e que me iluminam ainda mais os dias: o Vicente e o David.

Ao Paul Hardman. Acreditou em mim desde a primeira aula como meu professor até ao momento em que se tornou colega de trabalho e amigo. Inseriu-me na vida selvagem do mercado e apoiou sempre as aventuras a que me propus neste trabalho.

Ao João Cunha. Deu-me sempre direcção, nunca me deixando desviar do caminho e facilitando sempre as partes menos agradáveis deste trabalho.

Ao Rafael Ferreira. Desde cedo no curso que se tornou no meu parceiro de guerra em qualquer que fosse o trabalho. Foi também a pessoa com quem mais aprendi na nossa área. Foi fácil manter essa parceria até aos dias de hoje. Foi fácil transformá-la em amizade. Hoje podemos dizer que juntos temos um nome: —, *you had a problem*.

Aos meus amigos de curso (colegas não lhes assenta bem). O meu percurso nunca teria seguido o caminho que seguiu sem eles. É difícil explicar o espírito de entreaajuda que existiu sempre ao longo desta aventura. A colaboração em vez da competição. Levo essa forma de trabalhar para a minha vida. Mas, acima de tudo, levo excelentes amigos: Filipa Pereira, Jéssica Mendes, Jéssica Parente, Adriana Nunes, Mariana Monteiro, Sérgio Rebelo e Pedro Silva.

Ao Duarte Emi. Coimbra ganhou outra vida com a sua presença. Amigo desde o início e capaz de tornar cada dia ainda mais imprevisível. Foi da música que nasceu esta amizade e é com ela que continua a crescer.

Aos professores que fizeram parte da minha formação. Com todos aprendi, nas boas e nas mais difíceis experiências. Fizeram parte do meu crescimento como designer multimédia, mas também como pessoa.

Ao Adriano Esteves. Foi uma peça fulcral na fase final deste projecto, devido a toda a compreensão e apoio que tem prestado e a tudo o que, com ele, tenho aprendido.

A todos, o meu enorme agradecimento, por me terem tornado no designer multimédia, no músico e, acima de tudo, na pessoa que sou.

Resumo

—

Hoje em dia, os álbuns de música são, muitas vezes, vistos como um conjunto de peças não relacionadas em vez de se apresentarem com uma identidade própria. Desde o início da indústria fonográfica houve diversas mudanças de paradigma no design e no marketing, que coincidiram com o surgimento e o domínio de novas tecnologias, estendendo-se também para a relação existente entre os álbuns de música e os concertos relacionados. Nos últimos 20 anos estas mudanças intensificaram-se de tal modo que se tornou complicado para os diversos integrantes da indústria fonográfica e da música ao vivo acompanhá-las.

Nesta dissertação propõe-se uma nova abordagem para a unificação de todos os materiais e eventos que um álbum de música abrange a partir de um só sistema, que relaciona as áreas do design, multimédia e música e com o qual se produz imagens estáticas e animadas baseadas nas letras e no som presente nas músicas de um álbum. Com o output estático gerado pelo sistema, é possível desenvolver uma imagem gráfica para o álbum e com o output animado, pode criar-se videoclipes e efeitos visuais para os concertos relativos ao mesmo álbum. Este sistema permite a coesão gráfica entre todos os materiais e eventos que integram um álbum, enaltecendo os dois elementos que se consideram mais importantes numa música: a letra, que é portadora da mensagem do artista, e o som, que transmite emoções à mensagem.

palavras-chave

—
design generativo
design de som
creative coding
ilustração generativa
análise de texto
Processing
análise de som
Pure Data
design gráfico
álbum de música
efeitos visuais

Abstract

—

Music albums are often seen as a set of unrelated pieces rather than presenting themselves with an identity of their own. Since the beginning of the music industry there have been several paradigm shifts in design and marketing, which have coincided with the emergence and prevalence of new technologies, extending to the relationship between music albums and related concerts. In the past 20 years these changes have intensified in such a way that it has become difficult for the various members of the music industry and live music to keep up with them.

This dissertation proposes a new approach for the unification of all materials and events associated to a music album out of a single system. This system creates a relationship between the areas of design, multimedia and music and produces static and animated images based on the lyrics and the sound present in the songs of an album. With the static output, it is possible to develop a graphic image for the album. The system's animated output is able to create video clips and visual effects for the concerts related to the same album. This system allows the graphic cohesion between all the materials and events that integrate an album, highlighting the two elements that are considered most important in a song: the lyrics, which carries the message of the artist, and the sound, which transmits emotions to the message.

keywords

—
generative design
sound design
creative coding
generative illustration
text analysis
Processing
sound analysis
Pure Data
graphic design
music album
visual effects

Índice

—

I	Introdução	15
	—	
1.1	Motivação	19
1.2	Âmbito	20
1.3	Enquadramento	22
1.4	Objectivos	23
1.5	Estrutura do documento	24
II	Contextualização Histórica	27
	—	
2.1	Breve história dos meios físicos de armazenamento de música	29
2.2	Efeitos da “pirataria” na indústria da música	30
2.3	Novos paradigmas da revolução digital	32
2.4	Contribuição das editoras discográficas no papel do design na música	42
2.5	Mudança de paradigma na relação concertos-álbuns de música	48
2.6	Festivais vs. Concertos individuais	49
2.7	Breve história dos efeitos visuais em concertos de música	49
2.8	Novos métodos de actuação e de interacção com o público	50
2.9	Conclusões	52
III	Casos de Estudo	55
	—	
3.1	Geração de imagens a partir de texto	57
3.2	Geração de som a partir de texto	60
3.3	Ilustração generativa	60
3.4	Modelação de imagem de acordo com som	61
3.5	Ferramentas de geração de capas	66
3.6	Álbuns de música	70
3.7	Conclusões	74
IV	Metodologia	77
	—	
4.1	Objectivos	79
4.2	Processo	80
4.3	Plano de trabalho	83

V	Projecto	87
	—	
5.1	Descrição do projecto	90
5.2	Módulo #1 - Análise e preparação da letra da música	92
5.3	Módulo #2 - Definição dos tempos das palavras-chave na música	98
5.4	Experiências preliminares de ilustração generativa	99
5.5	Módulo #3.1 - Recolha de imagens e aplicação do filtro <i>Canny Edge Detector</i>	103
5.6	Módulo #3.2 - Aprimoramento das imagens base	106
5.7	Módulo #4 - Ilustração e animação com base nas imagens filtradas	111
5.8	Módulo #5 - Mapeamento de atributos sonoros com variáveis visuais	115
5.9	Atributos sonoros estudados mas não utilizados	130
5.10	Módulo #6 - Possíveis aplicações do sistema num álbum de música	136
5.11	Módulo #7 - Aplicação do sistema nos efeitos visuais de um concerto	144
5.12	Discussão de resultados	146
VI	Conclusão	151
	—	
6.1	Trabalho futuro	155
VII	Referências Bibliográficas	157
	—	
7.1	Bibliografia	159
7.2	Webgrafia	161
	Anexo — Artigo e Poster ICCC 2018	171
	—	

Introdução

Capítulo I

- 1.1 Motivação
- 1.2 Âmbito
- 1.3 Enquadramento
- 1.4 Objectivos
- 1.5 Estrutura do documento

Capítulo I

Introdução



Esta dissertação pretende compreender o papel do design e da multimédia na música, explorando a questão: Como pode um só sistema unificar um álbum de música em todos os seus materiais estáticos e dinâmicos, e até eventos, num só vocabulário visual, enaltecendo os aspectos que se consideram mais importantes numa música — a mensagem da letra e as emoções contidas no som — de modo a contribuir para uma melhor interpretação?

Nos últimos 20 anos o universo da música tem sofrido grandes mudanças devido a vários factores. A crescente facilidade de acesso a computadores pessoais, a migração destes computadores para outros tipos de dispositivos móveis, o crescimento da Internet, o surgimento das redes sociais e o aparecimento do *streaming* são alguns dos principais fenómenos que tiveram impacto na evolução da música. Não tem sido fácil para artistas, editoras, rádios e outros profissionais ligados à área, acompanhar todas estas mudanças.

A possibilidade de gravar, misturar e masterizar músicas no próprio computador em casa, com qualidade aceitável sem ter de recorrer a um estúdio profissional, fez com que o número de artistas crescesse de forma rápida. Além disso, a Internet, mais propriamente as redes sociais, permitiu que estes artistas se promovessem sem terem de recorrer a editoras ou a rádios.

O surgimento do *streaming* fez com que todo o negócio da música sofresse uma significativa revolução, levando a que um grande número de pessoas deixasse de comprar discos e de ouvir rádio para ouvir música nestas plataformas, sem custos e com possibilidade de escolha. Por outro lado, permitiu que mais artistas alcançassem uma maior quantidade de pessoas e tenham possibilidade de ganhar popularidade, sem terem de recorrer a editoras ou a rádios.

No entanto, existem hoje, por diversos motivos, muitos mais *one hit wonders* — artistas que alcançam a fama apenas com um single de sucesso, voltando, posteriormente, à obscuridade. As plataformas de *streaming* suscitaram um crescimento das chamadas *playlists* (que inserem músicas isoladas, perdendo o seu contexto no álbum em que cada uma se insere). Além disso, muitos artistas deixaram de trabalhar sob o conceito de álbum e passaram a trabalhar apenas com *singles* lançados isoladamente nas plataformas online.

Um álbum de música é um conjunto de músicas integradas numa determinada narrativa que pode existir na sonoridade, nos conteúdos das letras, na imagem gráfica ou simplesmente na cronologia da criação das

faixas. Todos estes fenómenos estão a levar a que os artistas que ainda defendem o conceito de álbum e dos objectos físicos a ele associados pensem em formas diferentes de fazer com que a sua música chegue às pessoas e as faça comprá-la. Um fenómeno que ajudou muito neste processo foi o crescimento das vendas dos vinis desde 2007¹, com especial destaque para o ano de 2016, que foi o ano de maior crescimento das suas vendas desde 1991². Muitos artistas passaram a investir em álbuns em formatos físicos para colecção, com estratégias de imagem gráfica interactiva que colocam o ouvinte numa posição activa e o inserem, de alguma forma, na mensagem presente no álbum. Alguns artistas têm investido em álbuns estritamente digitais, em diversos formatos, tais como *websites* interactivos, *visual albums*, etc.. Outros artistas, ainda, optam por desenvolver imagens gráficas que unem tudo o que está relacionado com o álbum, tanto na componente física como na digital: objectos físicos (CD, vinil, etc.), diferentes materiais de promoção de concertos, materiais de divulgação do álbum, efeitos visuais dos concertos, videoclipes, etc..

O foco desta dissertação baseia-se na relação entre texto, som e imagem. A ideia de relacionar o texto e som de uma música com a imagem, deu origem ao desenvolvimento de um sistema de geração de imagens que se baseiam em letras de músicas e que são modeladas de acordo com o som. Isto faz com que exista uma ligação directa entre os três componentes.

São exploradas formas de gerar ilustrações (imagem) a partir de letras de músicas (texto) e modelá-las de acordo com características sonoras da música (som), com o objectivo de demonstrar visualmente as mensagens e emoções da música. Estruturam-se estas tarefas num sistema cujo *output* pode ser utilizado em formato estático, como, por exemplo, numa imagem gráfica de um álbum, ou dinâmico, como, por exemplo, num videoclipe ou nos efeitos visuais de um concerto. Todo este processo tem como finalidade unir, a partir de um único sistema, todos os objectos e eventos associados a um álbum de música, atribuindo uma identidade gráfica que transpareça as suas mensagens e emoções. Pretende-se que esta abordagem sustente o conceito de álbum como um todo que abrange todos os materiais e eventos associados a um conjunto de músicas.

1 Statista. <https://www.statista.com/chart/1465/vinyl-lp-sales-in-the-us> (data de consulta: 16.12.2017)

2 The Guardian. <https://www.theguardian.com/music/2017/jan/03/record-sales-vinyl-hits-25-year-high-and-outstrips-streaming> (data de consulta: 16.12.2017)

1.1 **Motivação**

A música, para além de ser uma forma de expressão artística é também um veículo de transmissão de mensagem, através da letra e das emoções que desencadeia.

As formas de transmissão desta mensagem, nomeadamente os álbuns de música — que agregam todas as mensagens num conceito comum que origina um objecto artístico — e os concertos — que podem ser encarados como uma conversa entre o artista e o seu público — são dignos de um estudo aprofundado. Os músicos têm necessidade de definir estratégias para fazer chegar ao público a música que fazem e a mensagem que pretendem transmitir. A inovação é uma estratégia incontornável neste domínio. O recurso à tecnologia tem sido, desde sempre, crucial na resposta a esta necessidade. Assim, nesta dissertação pretende-se **i)** pesquisar sobre o que tem sido feito com sucesso na área da divulgação da música e **ii)** explorar técnicas nas áreas do design e da multimédia que permitam novas abordagens que apelem à atenção do público para a música.

Recorre-se à exploração do design gráfico a partir do aprofundamento de conhecimentos e competências na área da programação criativa para criar um sistema que gere imagens dependentes da letra e do som da música, permitindo uma melhor interpretação dessa música.

Na pesquisa feita, não foram encontrados projectos ou estudos com o mesmo objectivo e a mesma abordagem. É neste quadro que se insere a motivação para este projecto.

1.2 Âmbito

Nesta dissertação, partindo de algumas referências acerca da evolução histórica da indústria da música (álbuns, concertos e outros suportes) até à actualidade, são identificados alguns problemas associados à necessidade que os músicos têm de conquistar o seu público. São ainda abordadas algumas soluções para esses problemas, bem como as dificuldades que hoje se sentem com a divulgação em massa de música na Internet. É neste contexto histórico que se enquadra o objectivo geral desta dissertação — recuperar o conceito de álbum como mensagem e objecto artístico, enaltecendo a mensagem contida na letra e a emoção desencadeada pelo som da música.

A componente prática desta dissertação assenta numa estratégia que pressupõe a relação da imagem com o som da música e com o texto da mensagem. Com este propósito, pretende-se desenvolver um sistema de geração de imagens a partir do texto das letras das músicas, sendo modeladas de acordo com características do som. Pretende-se que o *output* do sistema se constitua como um recurso que permite criar uma imagem gráfica para um álbum de música, sendo esta abordagem a base de sustentação do conceito de álbum, integrando todos os materiais e eventos associados à sua divulgação.

Para o desenvolvimento desta componente prática foi necessário efectuar uma pesquisa sobre a situação actual no que respeita à geração de imagens a partir de texto, geração de som a partir de texto, ilustração generativa, modelação de imagem de acordo com o som, ferramentas de geração de capas, exemplos de imagens gráficas de álbuns de música e de efeitos visuais de concertos que lhes correspondem.

O que se pretende não é criar uma imagem gráfica para um álbum ou para um concerto, mas sim uma ferramenta que permita uma interação entre som/texto e a imagem de modo a criar um conceito para um álbum no sentido de ser uma colecção de materiais/eventos associados a um conjunto de músicas.

Apesar de se abordar a temática do marketing no estado da arte, só é importante para esta dissertação na medida em que a vertente criativa contribui para a divulgação da música, tendo este trabalho esse mesmo objectivo. Não se pretende, de modo algum, um foco na vertente comercial do marketing.

Espera-se criar um novo conceito de álbum que recupere o antigo, que reúne um objecto artístico e uma mensagem, mas utilizando novos instrumentos tecnológicos, permitindo, assim, abrir novas portas. Deste modo, poderá existir uma união do aspecto físico e analógico do álbum clássico com a componente digital e tecnológica presente nesta nova era.

1.3 Enquadramento

Esta dissertação insere-se no Mestrado em Design e Multimédia da Universidade de Coimbra, que privilegia a integração de duas áreas: uma mais criativa, a do design, e outra mais técnica, a da utilização de ferramentas digitais.

O tema abordado na dissertação — a relação música-mensagem-imagem gráfica — tem conhecido uma evolução interessante desde que foi possível gravar música pela primeira vez, em finais do século XIX, até aos nossos dias. No entanto, alterações muito significativas, associadas a uma revolução digital no mundo da música, trazem novos paradigmas. Apesar dos inúmeros aspectos positivos, estas mudanças levantam alguns problemas que têm posto em causa a integridade dos álbuns produzidos pelos músicos. A componente teórica desta dissertação situa em termos históricos esta problemática. Na componente prática, desenvolve-se um sistema que gera um *output* que se baseia na mensagem das letras e nas características sonoras das músicas, com recurso a ilustração generativa modelada por som, que pode ser utilizado para criar uma imagem gráfica para um álbum e todos os eventos/materiais a ele associados. Na pesquisa efectuada, não se encontrou nenhuma abordagem com o mesmo objectivo.

Em tempos de supremacia da tecnologia, acredita-se que trazer esta nova abordagem à relação texto/som e imagem possa oferecer mais significado à música e recuperar o conceito de identidade de um álbum, relacionando-o com os respectivos concertos. Além de se pretender contribuir para a inovação, sempre necessária no mundo da música, espera-se ainda oferecer um contributo para o desenvolvimento da integração do design e multimédia, sobretudo nos domínios do design generativo, design de interacção, design de som e design gráfico.

1.4 Objectivos

O principal objectivo desta dissertação é encontrar uma forma de proporcionar uma interpretação visual focada nos aspectos que se consideram mais importantes numa música: letra, portadora da mensagem, e som, meio de transmissão de emoções. Por outro lado, também se pretende recorrer a essa interpretação visual para criar uma imagem gráfica de um álbum de música, de modo a unificar todos os diferentes materiais e eventos a ele associados. Desta forma, acredita-se que seja possível recuperar o conceito de álbum como mensagem e objecto artístico.

Estes objectivos implicam uma exploração e um estudo aprofundados, quer do conhecimento neste domínio, quer de ferramentas com as quais é possível tratar computacionalmente texto, imagem e som, de modo a poder relacionar estas componentes.

1.5 Estrutura do documento

Este documento está dividido em sete capítulos: **Introdução, Contextualização Histórica, Casos de Estudo, Metodologia, Projecto, Conclusão e Referências bibliográficas**. Apresenta-se o **Anexo — Artigo e Poster ICCC 2018**.

Na **Introdução**, apresenta-se uma breve contextualização que situa a dissertação nas diferentes áreas abordadas e suas problemáticas, para posteriormente se descrever a ideia geral que está na base do projecto, os passos principais do processo e a apresentação da questão a explorar. Além disso, apresentam-se **i)** os factores que motivaram à elaboração deste projecto, **ii)** o âmbito em que se insere, esclarecendo o que vai e não vai ser abordado, **iii)** de que forma se insere no contexto actual, **iv)** os objectivos que se pretende alcançar e, por fim, **v)** a estrutura sob a qual foi elaborado o documento que apresenta as componentes teórica e prática da presente dissertação.

O capítulo **Contextualização Histórica** está dividido em diversas secções que vão desde o surgimento da indústria fonográfica até aos paradigmas actuais impostos pela revolução digital. Posteriormente, estabelece-se um paralelo desta narrativa com um pedaço da história da indústria da música ao vivo, demonstrando a sua ligação e as flutuações na influência de cada uma na outra. Desta forma, situa-se a dissertação no contexto histórico e actual das duas indústrias, clarificando o propósito do projecto de criar um sistema que sirva de união visual entre o álbum de música e os concertos correspondentes.

No capítulo seguinte, **Casos de Estudo**, são apresentados diversos projectos actuais que servem de referência a cada etapa presente na componente prática da dissertação. Assim, cria-se uma divisão do conjunto de projectos seleccionados por seis secções, referentes a essas etapas: geração de imagens a partir de texto, geração de som a partir de texto, ilustração generativa, modelação de imagem de acordo com som, ferramentas de geração de capas e álbuns de música. Assim, é possível criar um panorama do que tem sido feito actualmente em cada área abordada na dissertação, apresentando-se as principais fontes de inspiração para este projecto, identificando-se alguns casos que apresentavam fragilidades que se tentou evitar ao longo do projecto.

No capítulo **Metodologia**, enumeram-se os objectivos específicos, descreve-se globalmente o processo de execução do projecto e apresenta-se o plano de trabalho que foi cumprido.

No capítulo **Projecto**, depois de uma introdução, passa-se à descrição detalhada das componentes estruturantes do sistema desenvolvido na componente prática do projecto. Segue-se depois a descrição dos diferentes módulos que integram o referido sistema. O capítulo termina com uma secção onde se discutem os resultados obtidos, apresentando algumas possibilidades de exploração do sistema desenvolvido no futuro.

No capítulo **Conclusão**, encerra-se o documento com um breve resumo das ideias fundamentais presentes nesta dissertação e com que se pensa ter contribuído para o conhecimento e a inovação na área do Design e Multimédia, apontando algumas linhas de exploração para trabalho futuro.

Por fim, surge um **Anexo — Artigo e Poster ICC 2018** —, no qual se apresenta o artigo relativo a este projecto que foi integrado na publicação *Proceedings of the 9th International Conference on Computational Creativity*, juntamente com o poster que foi apresentado na conferência.

Contextualização Histórica

Capítulo II

- 2.1 Breve história dos meios físicos de armazenamento de música
- 2.2 Efeitos da “pirataria” na indústria da música
- 2.3 Novos paradigmas da revolução digital
- 2.4 Contribuição das editoras discográficas no papel do design na música
- 2.5 Mudança de paradigma na relação concertos-álbuns de música
- 2.6 Festivais vs. Concertos individuais
- 2.7 Breve história dos efeitos visuais em concertos de música
- 2.8 Novos métodos de actuação e de interacção com o público
- 2.9 Conclusões

Capítulo II

Contextualização Histórica

—

Esta dissertação assenta no conceito de álbum de música como a união do objecto artístico com todos os materiais e eventos associados, com o objectivo de transmitir mensagens e emoções ao seu público. Para levar a cabo esta complexa tarefa, é necessário conhecer o contexto histórico e o da actualidade em que este conceito se enquadra.

Na primeira parte deste capítulo, descreve-se a história da indústria fonográfica na perspectiva da evolução dos paradigmas do design e do marketing de acordo com o surgimento e o domínio de novas tecnologias ao longo dos anos. Na segunda parte, estabelece-se um paralelo entre esta narrativa e a história da indústria da música ao vivo, com o foco presente na mudança de paradigmas da relação entre os álbuns de música e os concertos.

Desta forma, insere-se esta dissertação nas indústrias fonográfica e da música ao vivo, para se justificar o seu funcionamento como conector entre elas através do design e da multimédia, a fim de promover o álbum como objecto artístico e como conjunto de materiais e eventos associados.

2.1 Breve história dos meios físicos de armazenamento de música

Ouvir música, até meados do séc. XIX, era uma experiência que apenas era possível em actuações ao vivo. Só em 1877, graças à invenção do fonógrafo e dos seus cilindros fonográficos por Thomas Edison, se tornou possível comercializar música e, como tal, passar a ouvir música sem ser apenas a partir de concertos³. A partir daí, novas tecnologias foram surgindo, tornando as antigas obsoletas e impondo novos paradigmas na indústria da música. A partir do fonógrafo surgiu o gramofone, por Emile Berliner, que substituiu os cilindros fonográficos por discos de 78 rpm, mais resistentes e mais práticos⁴. Estes discos mantiveram o seu domínio até 1948, ano em que a Columbia Records introduziu o LP de 33 $\frac{1}{3}$ rpm (*long-play record*) no mercado e que, dentro de um ano, já era a principal escolha do público e da indústria (Jones & Sorger, 1999, p. 74). Ao mesmo tempo, em 1949, a RCA Victor lançou o disco de 45 rpm, que competia com o LP pelo seu tamanho

3 Library of Congress. <https://www.loc.gov/collections/edison-company-motion-pictures-and-sound-recordings/articles-and-essays/history-of-edison-sound-recordings/history-of-the-cylinder-phonograph/> (data de consulta: 02.01.2018)

4 Library of Congress. <https://www.loc.gov/collections/emile-berliner/articles-and-essays/gramophone/> (data de consulta: 02.01.2018)

menor e, conseqüentemente, custo mais acessível⁵. Durante os dez anos seguintes, os discos de 45 rpm serviam principalmente para gravar *singles* ou *extended-plays*, enquanto os LPs continham mais obras de jazz ou música clássica, geralmente adquiridos por colecionadores (Jones & Sorger, 1999, p. 74). Em meados dos anos 1960, o LP passou ao topo do mercado devido ao sucesso do *rock and roll*, o que veio trazer segurança às bandas para gravar álbuns completos em vez de êxitos isolados (*Ibidem*, p. 76). Esta posição do LP no mercado ficou ainda mais consolidada com o aparecimento de novas formas de *rock* nos anos 1970. Entretanto, a Phillips introduziu a fita cassete no mercado em 1963, que permitiu não só reproduzir mas também gravar conteúdo (Tschmuck, 2012, p. 163), levando a que esta fosse a primeira forma conhecida da chamada “pirataria” (*Ibidem*, p. 164). Com a invenção do *Walkman* por parte da Sony, em 1979 (*Ibidem*, p. 166), a fita cassete entrou em verdadeira disputa com o LP e acabou mesmo por ultrapassá-lo nas vendas em 1983⁶. O LP foi perdendo força no mercado, não só por causa da fita cassete, mas também por causa da introdução do CD (*compact disk*) no mercado em 1983. Foi então que arrancou a revolução digital, pois passou a ser possível armazenar informação digital de forma comprimida num formato mais fácil de manusear e transportar do que o LP e sem os ruídos típicos dos formatos analógicos (Tschmuck, 2012, p. 166). A “pirataria” ganhou mais força, pois agora era possível gravar músicas em dois formatos diferentes. Em 1992, o CD ultrapassou a cassete nas vendas e manteve o seu crescimento até 2000, ano a partir do qual começou a drástica queda nas vendas dos formatos físicos⁷.

2.2 Efeitos da “pirataria” na indústria da música

O debate à volta da “pirataria” nunca foi, até hoje, concluído e dividiu a indústria da música, os artistas e os consumidores em todo o mundo. No início dos anos 1990, a Internet começou a espalhar-se por um número cada vez maior de computadores e, com a introdução do *M-Bone-Systems* (que permitia a transmissão de sinais áudio e vídeo) e do formato digital de áudio comprimido chamado *MP3*, transformou-se também num canal de distribuição de música.

Em 1998, surgiu o Napster, um software que permitia que os seus utilizadores partilhassem músicas que estivessem armazenadas nos seus computadores (Tschmuck, 2012, p. 182-183). Foi aqui que a “pirataria” ganhou a sua máxima força e a própria indústria da música começou a tentar combater a Napster. No entanto, este novo paradigma já estava consistentemente instalado (começaram a surgir diversos softwares semelhantes) e não havia como eliminá-lo definitivamente. Bandas como Beastie Boys, Dave Matthews, Moby e Limp Bizkit apelavam à utilização do Napster, enquanto Peter Gabriel, Eminem, Metallica e Dr. Dre alegavam

5 CBS News. <https://www.cbsnews.com/news/almanac-the-45-rpm-record/> (data da consulta: 6.01.2018)

6 RIAA. <https://www.riaa.com/u-s-sales-database/> (data da consulta: 6.01.2018)

7 *Idem*.

violação de direitos de autor e danos económicos (*Ibidem*, p. 184). Peter Tschmuck analisou diversos estudos relativos a este assunto no seu livro *Creativity and Innovation in the Music Industry* e dividiu-os em dois grupos:

Efeito de substituição — estudos que defendem que a partilha de ficheiros veio substituir a venda de música, provocando, assim, um impacto negativo.

Efeito de *sampling*, exposição ou penetração — estudos que defendem que a partilha de ficheiros facilita a exposição dos artistas, levando até a um impacto positivo nas vendas⁸.

Peter Tschmuck apresenta, após a análise destes estudos, uma nova teoria para o declínio nas vendas que tem havido desde 2000 e, para a explicar, retorna a 1970 (Gráfico 1), quando o mercado se dedicou principalmente aos formatos *long-play*, assinalando um diferente paradigma: os *singles* passaram a servir para testar um artista ou banda no mercado, e só se estes fossem bem sucedidos é que se avançava para um álbum. No entanto, o álbum tem a desvantagem de, em geral, apenas conter uma ou outra faixa que realmente interessam ao consumidor, e quando este paradigma foi confrontado com o da Internet — que possibilitou consumir faixas isoladas (seja por “pirataria” ou não) — os consumidores optaram por investir apenas nas faixas que lhes interessavam. Assim, conclui que a “pirataria” não é uma causa da descida das vendas, mas sim um sintoma da revolução digital na indústria musical (Tschmuck, 2012, p. 189-190).

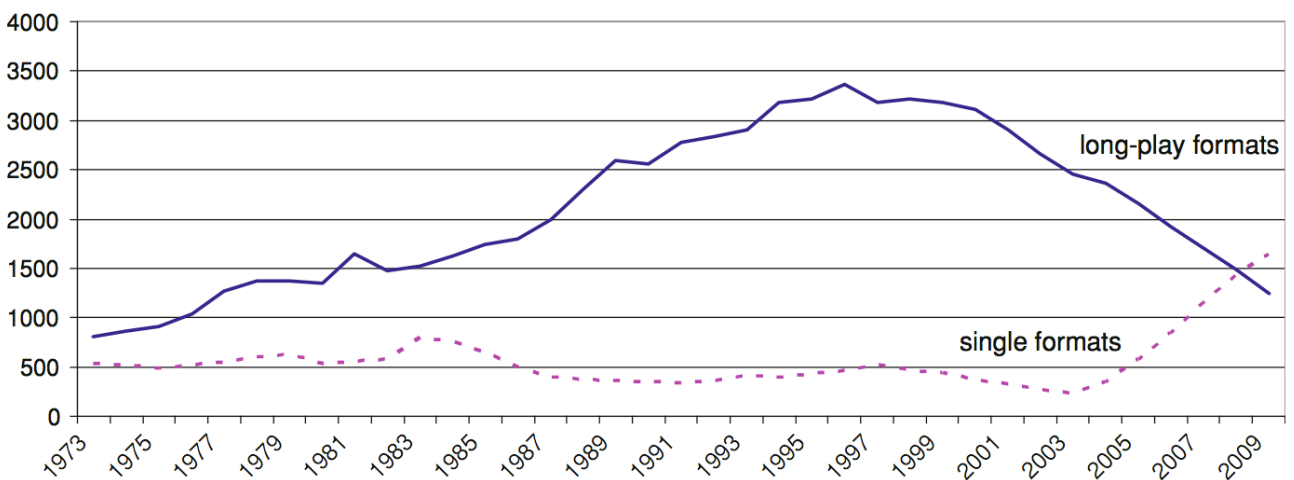


Gráfico 1 — Fonte: IFPI (1973—2010). As evoluções dos *singles* e dos *long-plays* comparados. Screenshot do artigo.

⁸ cf. com *Creativity and Innovation in the Music Industry*, Peter Tschmuck na pág. 188.

2.3 Novos paradigmas da revolução digital

Enquanto se mantinha o aceso debate sobre a “pirataria”, a tecnologia avançava e novos modos de distribuição e marketing foram surgindo em intervalos de tempo muito curtos. Houve várias tentativas de lançar plataformas online para venda de música, no entanto, foi a Apple que conseguiu ser bem sucedida no mercado, quando lançou a iTunes Music Store, em 2003, e o volume de vendas disparou. Nos anos seguintes, várias plataformas entraram também na venda online de música, nomeadamente a Amazon e a Google (*Ibidem*, p. 191), levando a que as subidas das vendas online de ano para ano continuassem até 2012. Desde 2012, no entanto, tem havido uma descida das vendas online⁹, que coincide com o crescimento da popularidade das plataformas de *streaming*, principalmente o Spotify, que se sustentam com publicidade e subscrições pagas. À conta da popularidade do *streaming*, instalou-se um novo paradigma: tornou-se mais importante o acesso à música do que a sua posse (*Ibidem*, p. 192). Este paradigma obrigou a indústria e os artistas a alterar completamente a sua estratégia de marketing, para conseguirem contornar os problemas que a revolução digital trouxe às vendas.

Alguns artistas recusaram-se a disponibilizar a sua música neste tipo de plataforma, por não concordarem com o paradigma do *streaming*. Taylor Swift, uma das cantoras mais famosas do mundo, retirou todas as suas músicas do Spotify em 2014 (embora as tenha mantido em outras plataformas, como o TIDAL), alegando que o retorno não reflectia o verdadeiro valor da arte que é a música¹⁰. No entanto, em Junho de 2017, voltou a disponibilizar todos os seus álbuns no Spotify, possivelmente, devido ao facto de a sua dimensão e importância serem demasiado elevadas nos dias de hoje para ignorar¹¹.

Outros artistas optaram por não disponibilizar as suas músicas em nenhuma plataforma de *streaming*¹², sendo um deles o Thom Yorke, o vocalista dos Radiohead, dos Atoms For Peace e com trabalho a solo. Yorke falou, numa entrevista ao *website* mexicano *Sopitas*¹³, da experiência que fez em 2007 com os Radiohead no lançamento do álbum *In Rainbows* (Fig. 1) — a própria banda (sem qualquer editora) lançou o álbum no seu *website*, disponibilizando-o para *download* com a estratégia *pay-what-you-want*¹⁴ — e explicou a razão pela qual o removeu do Spotify:

When we did the In Rainbows thing what was most exciting was the idea you could have a direct connection between you

9 RIAA. <https://www.riaa.com/u-s-sales-database/> (data da consulta: 8.01.2018)

10 Business Insider. <http://www.businessinsider.com/taylor-swift-explains-why-she-left-spotify-2014-11> (data de consulta: 9.01.2018)

11 Forbes. <https://www.forbes.com/sites/hughmcintyre/2017/06/27/why-did-taylor-swift-really-rejoin-spotify/#7d66446373de> (data de consulta: 10.01.2018)

12 cf. com Rolling Stone. <https://www.rollingstone.com/music/news/artists-refuse-stream-music-20150102> (data de consulta: 10.01.2018)

13 The Guardian. <https://www.theguardian.com/technology/2013/oct/07/spotify-thom-yorke-dying-corpse> (data de consulta: 10.01.2018)

14 Time. <http://content.time.com/time/arts/article/0,8599,1666973,00.html> (data de consulta: 10.01.2018)

as a musician and your audience. You cut all of it out, it's just that and that. And then all these fuckers get in a way, like Spotify suddenly trying to become the gatekeepers to the whole process.

Entretanto, o álbum já foi disponibilizado no Spotify no final de 2017, no entanto, Yorke continua a mostrar-se crítico relativamente à plataforma, tendo já partilhado um *tweet* de Geoff Barrow, instrumentalista nos Portishead e membro das bandas Beak e Quakers, no qual questiona quantos artistas fizeram mais de 500 libras com o Spotify¹⁵. Uma das respostas veio do produtor Dan Le Sac, “que alega ter recebido pouco mais de 100 libras (112 euros) por 20 mil escutas no Spotify”¹⁶.

A experiência *pay-what-you-want* dos Radiohead foi inovadora na indústria musical, pois ainda não tinha sido realizada por nenhuma banda ou artista mundialmente conhecidos. Os resultados foram revelados um ano depois e mostraram que a experiência foi um sucesso, apesar de mais pessoas

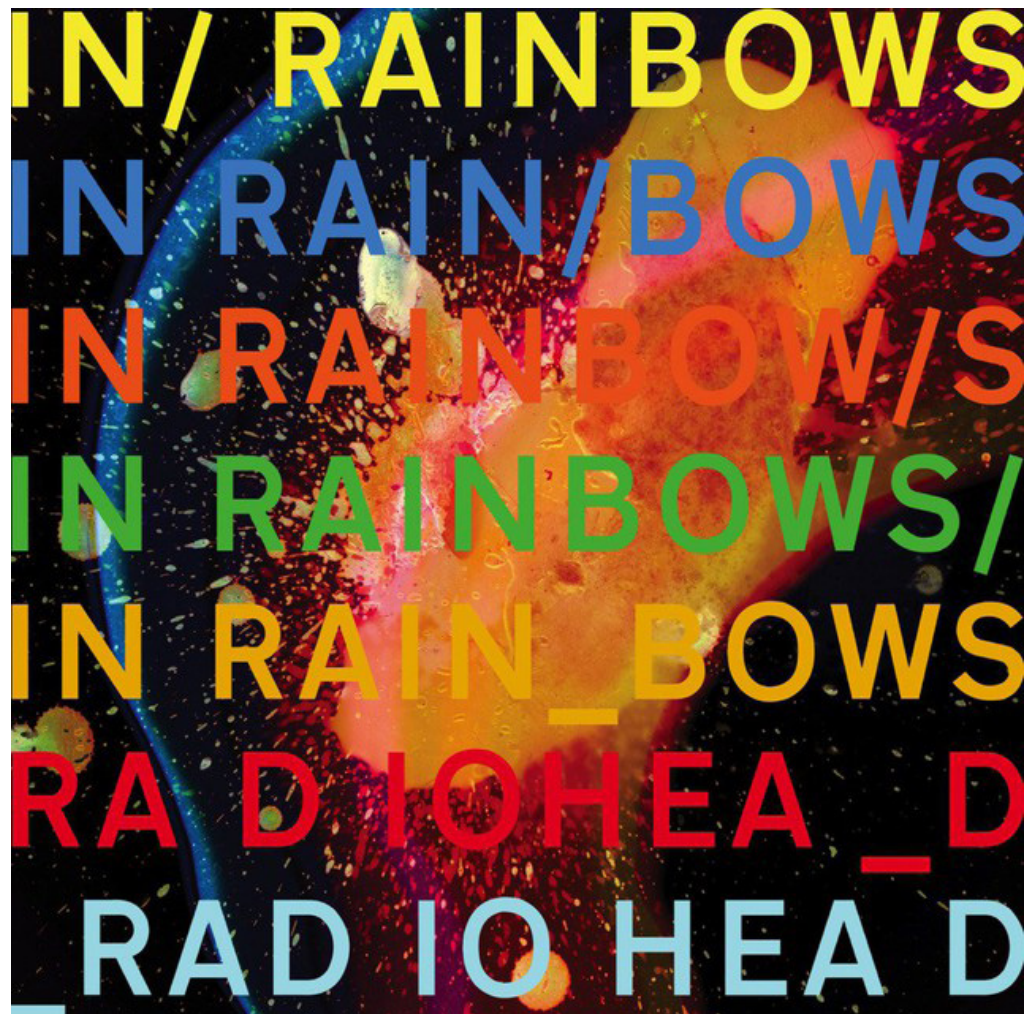


Figura 1 — Capa do disco *In Rainbows* dos Radiohead.

15 Twitter. <https://twitter.com/jetfury/status/946051057553113088> (data de consulta: 10.01.2018)

16 Blitz. <http://blitz.sapo.pt/principal/update/2017-12-29-Os-Radiohead-ja-tem-a-musica-no-Spotify-mas-Thom-Yorke-continua-zangado> (data de consulta: 10.01.2018)

terem adquirido o álbum gratuitamente do que a pagar: “Warner Chappell confirmed that ‘Radiohead made more money before *In Rainbows* was physically released than they made in total on the previous album *Hail To the Thief*”. O álbum físico foi lançado três meses depois (com músicas que não estavam na versão do álbum disponível para *download*) em CD, vinil e numa edição especial que vendeu cerca de 100 mil cópias que custavam 80 dólares cada uma, o que traduz um número de sucesso, tendo em conta a estratégia que permitia adquirir o álbum digital gratuitamente e o facto de ter havido mais *downloads* via BitTorrent do que no *website* dos Radiohead¹⁷.

Estes resultados demonstram claramente um dos paradigmas que domina hoje a indústria musical: utilizar os álbuns online para promover edições especiais dos álbuns em formatos físicos. E o que é certo é que as vendas dos vinis têm subido consideravelmente nos últimos anos. Segundo um artigo do jornal *The Guardian* lançado no início de 2017¹⁸, a morte de gigantes do mundo da música, como David Bowie, tem projectado grandes subidas nas vendas de vinis, pois as pessoas compram-nos como forma de recordação: David Bowie foi o artista com mais vinis vendidos em 2016, ano da sua morte. O início do crescimento nas vendas de vinis, em 2007, ganhou expressão também com o surgimento do evento *Record Store Day*, em 2008, que se tornou um fenómeno mundial e tem vindo a crescer de ano para ano. Ainda no artigo do jornal *The Guardian* acima mencionado, são citadas algumas palavras de Vanessa Higgins, CEO da *Regent Street* e da *Gold Bar Records* e membro representativo das editoras independentes no *BPI Council* (*British Phonographic Industry — Limited*):

People think millennials just stream and are just digital but actually I think we are going to see increasingly over this coming year that young people still want something tangible and real and that’s where vinyl is taking on the role that the CD used to have.

Higgins diz ainda que o crescimento do *streaming* — paradigma de ter acesso à música, mas não ser seu possuidor — levou a que as pessoas voltassem a comprar vinis como forma de possuir fisicamente a música: “They are finding music through streaming and if they love it, they are going out and investing in it in a physical format”.

À conta disto, muitos artistas têm explorado formas de *packaging* inovadoras para os seus vinis, retornando um pouco ao que já se fez nos anos áureos do LP, entre finais de 1960 e início de 1980.

17 Rolling Stone. <https://www.rollingstone.com/music/news/radiohead-publishers-reveal-in-rainbows-numbers-20081015> (data de consulta: 10.01.2018)

18 The Guardian. <https://www.theguardian.com/music/2017/jan/03/record-sales-vinyl-hits-25-year-high-and-outstrips-streaming> (data de consulta: 10.01.2018)

Alguns exemplos de álbuns dessa altura, que se tornaram icónicos, sobretudo devido ao *packaging*, são:

Sgt. Pepper's Lonely Hearts Club Band dos Beatles (Fig. 2), lançado em 1967 e desenhado por Peter Blake e Jann Howarth. Foi o primeiro álbum a ter as letras das músicas impressas e um cartão de recortes (Jones & Serger, 1999, p. 76), inspirando álbuns futuros a utilizar interacção para melhorar a experiência.



Figura 2 — Da esquerda para a direita: capa, contracapa com letras das músicas e interior para recortar.

School's Out de Alice Cooper (Fig. 3) foi desenhado pelos Pacific Eye & Ear e lançado em 1972. Tinha um *packaging* que abria de baixo para cima em vez da abertura normal de um livro, simulando uma secretária de escola. O vinil vinha embrulhado num par de cuecas feminino, ideia que teve de ser descontinuada devido ao facto de as cuecas serem inflamáveis, originando perigo no processo de fabrico¹⁹.



Figura 3 — Da esquerda para a direita: capa, interior a simular uma secretária de escola e o LP com o par de cuecas.

¹⁹ The Concourse. <https://theconcourse.deadspin.com/schools-out-forever-the-secret-history-of-alice-coop-1705441582> (data de consulta: 11.01.2018)

Sticky Fingers dos Rolling Stones (Fig. 4), desenhado por Andy Warhol e lançado em 1971, continha um *zipper* real que, ao ser aberto, revelava a roupa interior do modelo²⁰.

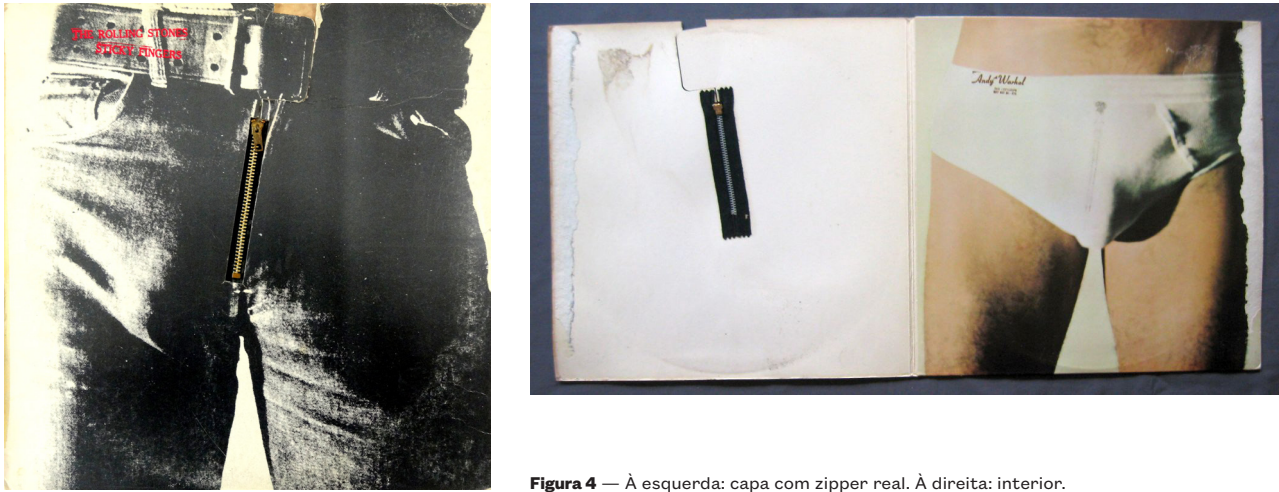


Figura 4 — À esquerda: capa com zipper real. À direita: interior.

Andy Warhol desenhou também o *The Velvet Underground & Nico* dos The Velvet Underground (Fig. 5), lançado em 1967, que tinha uma banana em autocolante que podia ser “descascada”, revelando no interior uma banana cor de rosa, capa esta que se tornou um ícone da *pop-art*²¹.

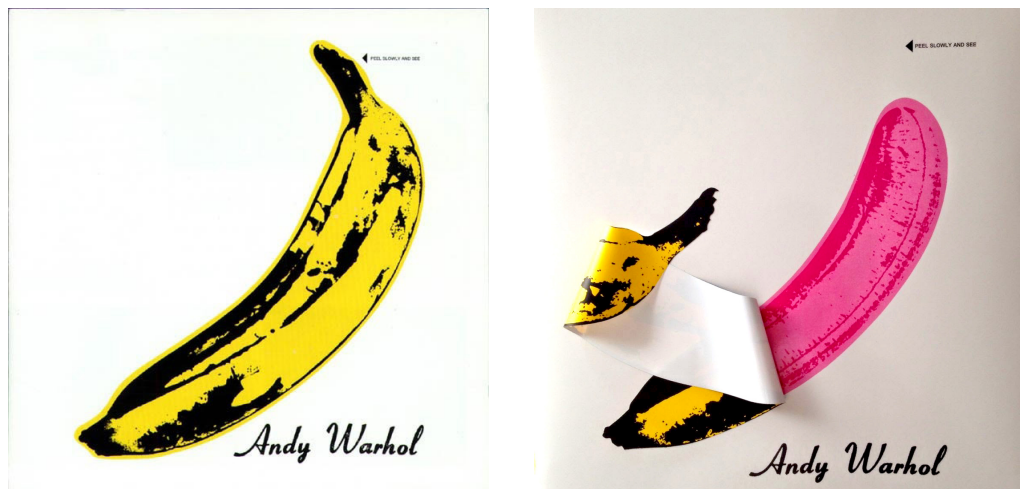


Figura 5 — À esquerda: capa com banana de autocolante. À direita: autocolante descolado, revelando o interior.

²⁰ Shutterstock. <https://www.shutterstock.com/blog/9-album-releases-that-turned-music-packaging-on-its-head> (data de consulta: 11.01.2018)

²¹ *Idem*.

Armed Forces de Elvis Costello And The Attractions (Fig. 6), desenhado por Barney Bubbles e lançado em 1979, que se tornou um ícone entre os *packagings* de álbuns devido à sua complexidade na construção: o consumidor tinha de passar por várias dobras com imagens diversificadas e muito coloridas para conseguir aceder ao disco, o que permitia uma interação e uma imersão num ambiente visual próprio para acompanhar a audição do álbum (Jones & Serger, 1999, p. 81).



Figura 6 — Abertura do *packaging* até chegar ao disco.

Entre os artistas que têm recuperado esta lógica, encontra-se Beck, que se destaca em dois dos seus álbuns por utilizar *packagings* interactivos para vender a sua música e reforçar o seu conceito: *The Information*, em 2006, e *Colors*, em 2017.



Figura 7 — Capa com grelha, autocolantes incluídos no interior do *packaging* e algumas capas de exemplo.

O primeiro (Fig. 7), desenhado pelos Big Active, é vendido com uma capa impressa apenas com uma grelha, contendo cada cópia no seu interior uma de quatro folhas diferentes com autocolantes com imagens de 20 artistas diferentes, num total de 250 imagens diferentes. O consumidor pode, assim, criar a sua própria capa com esses autocolantes e partilhá-la online com outros seguidores, ajudando, ao mesmo tempo, a promover o álbum²².

22 cf. com Big Active. <http://www.bigactive.com/art-direction/music/beck-information/> (data de consulta: 11.01.2018)

O segundo, *Colors* (Fig. 8), desenhado pelos designers Jimmy Turrell e Steve Stacey, tem também uma capa interactiva, mas apenas na edição *deluxe*, transparecendo o já mencionado paradigma actual de venda de álbuns físicos como objectos de colecção. O *packaging* contém diversas camadas coloridas recortadas em diferentes formas, podendo o consumidor dispô-las de diversos modos por detrás de um recorte hexagonal presente na capa, salientando o conceito do álbum. Turrell explicou um pouco desse conceito à *Dezeen* da seguinte forma²³:

It's music that makes you want to dance, is joyous and has been created specifically to be played out live. In a visual sense we wanted to run with that. So we went for punchy, geometric shapes and primary colours jammed together in a fun, almost childlike manner, but hopefully with a certain edge.

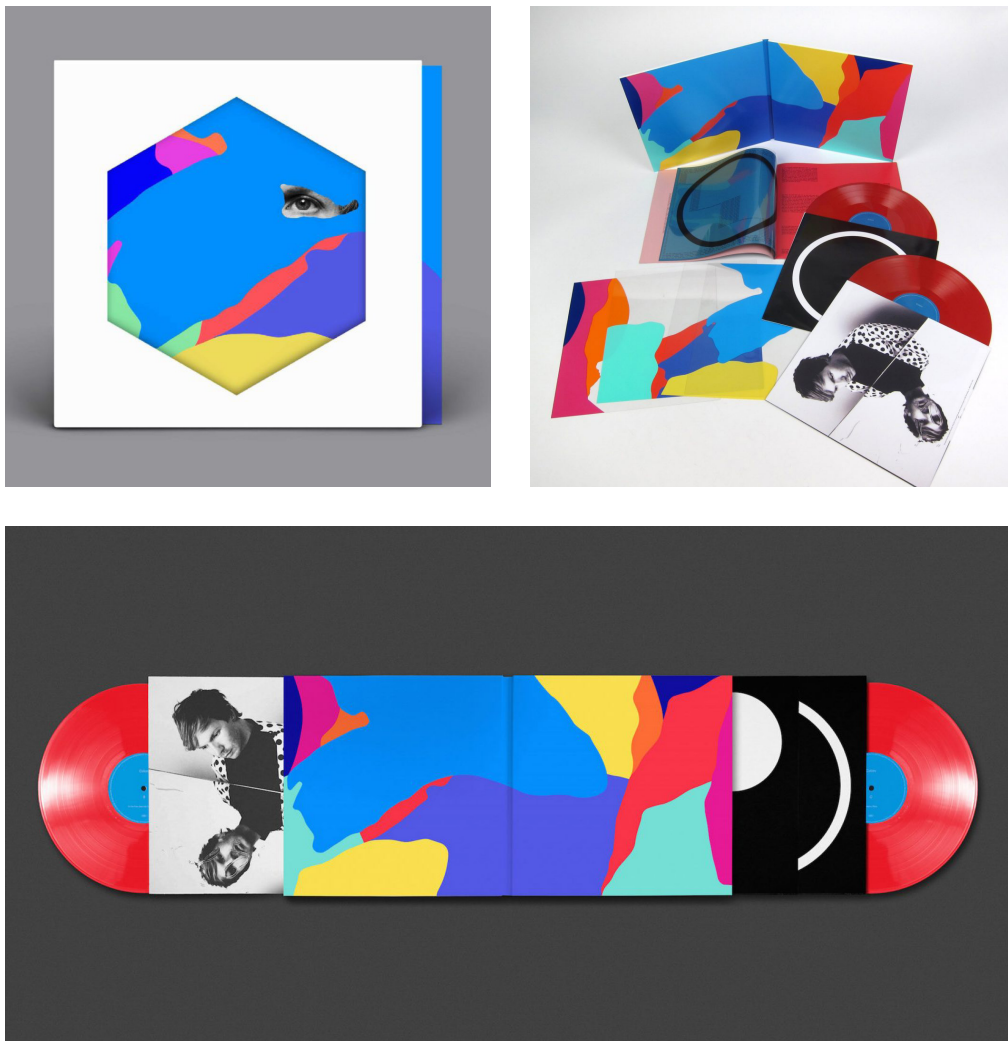


Figura 8 — Capa, restantes materiais e interior.

23 cf. com Dezeen. <https://www.dezeen.com/2017/09/30/beck-album-artwork-colours-customisable-design/> (data de consulta: 11.01.2018)

No entanto, o desafio dos artistas hoje em dia não se encontra apenas nos formatos físicos. Procuram encontrar formas de ser e de se manter relevantes online, no meio da imensidão de informação que existe na Internet.

Os Radiohead foram, uma vez mais, tema viral no mundo e na Internet. A *Buzz Magazine* fez um artigo com a descrição dos eventos que antecederam o lançamento do álbum *A Moon Shaped Pool* em 2016²⁴, aos quais acrescentarei também a minha experiência como seguidor da banda. No dia 30 de Abril desse ano, alguns seguidores receberam na sua caixa de correio um panfleto com a frase “Sing a song of sixpence that goes Burn the Witch.” e, imediatamente, expressaram a sua confusão nas suas contas de *Twitter*, *Facebook* e *Instagram*, tornando os Radiohead num dos maiores tópicos de conversa desse dia por todo o mundo: será que irão lançar um novo álbum? No dia 1 de Maio, apagaram todos os conteúdos do seu *website* e redes sociais, aumentando ainda mais a confusão, aparecendo nas notícias por todo o mundo e colocando ainda mais questões nos seguidores: será que a banda vai terminar? No dia 3 de Maio, durante a manhã, a banda lançou um pequeno *clip* na sua conta de *Instagram*, recentemente esvaziada, que mostrava uma animação de um pássaro feito de barro, que veio a revelar-se mais tarde como o início do videoclipe do *single Burn The Witch*. Mais tarde, no mesmo dia, lançaram outro *clip*, já com um pedaço do instrumental da música no fundo, que mostrava outra animação de barro de pessoas mascaradas a dançar à volta de uma mulher amarrada a uma árvore. Finalmente, a banda lançou o videoclipe do *single Burn The Witch* no *YouTube*, o primeiro *single* lançado em 5 anos. Lançaram mais um *single*, *Daydreaming*, no dia 6 e o álbum saiu no dia 8.

Esta estratégia de lançamentos-surpresa tem sido muito utilizada por diversos artistas mundialmente conhecidos desde o lançamento já descrito do álbum *In Rainbows* dos Radiohead²⁵. Outro exemplo representativo deste tipo de lançamento-surpresa foi o álbum homónimo de Beyoncé, lançado exclusivamente no *iTunes* no final de 2013 sem qualquer aviso prévio²⁶. Uma das grandes particularidades deste álbum foi o facto de Beyoncé ter feito um videoclipe para todas as faixas do álbum e os ter incluído também no *iTunes*, ao que ela chamou de *visual album*. Este conceito não é novo e já existe desde 1964, quando os Beatles lançaram *A Hard Day's Night*, um álbum acompanhado de um filme homónimo que integrava as suas músicas²⁷. Também os Pink Floyd, por exemplo, lançaram, em 1982, um filme homónimo do álbum *The Wall*, lançado em 1979, exactamente com o mesmo tipo de estrutura do *Lemonade* de Beyoncé, lançado em 2016. No entanto, o conceito de *visual album* tem mais impacto no mercado da era digital do que tinha antigamente, além de estar mais desenvolvido.

24 Buzz Magazine. <http://www.buzzmag.co.uk/uncategorized/moon-shaped-pool-radiohead-used-internet-disappear-completely-music-feature/> (data de consulta: 10.01.2018)

25 Fuse. <https://www.fuse.tv/2016/05/radiohead-in-rainbows-release-a-moon-shaped-pool> (data de consulta: 10.01.2018)

26 Billboard. <https://www.billboard.com/articles/columns/the-juice/5827398/beyonce-unexpectedly-releases-new-self-titled-visual-album-on> (data de consulta: 10.01.2018)

27 Cuepoint. <https://medium.com/cuepoint/the-history-and-evolution-of-visual-albums-77592e14304> (data de consulta: 10.01.2018)

Cara Harrison define *visual album* da seguinte forma: “The visual album is a hybrid medium between music video and film; like music video, it promotes an audio album, and like film, it is conceived as a whole work of art.” (Harrison, 2014, p. ii). A diferença entre estes videoclipes e os dos outros álbuns é que estes se completam e estão interligados numa só peça de arte, embora façam também sentido quando isolados (*Ibidem*, p. 9). Num dos cinco vídeos de apresentação do conceito do álbum no seu canal de *YouTube*, Beyoncé diz o seguinte²⁸:

I wanted people to hear things differently and have a different first impression. Not just listen to a 10 second clip but actually be able to see the whole vision of the album. It was important that we made this a movie, we made this an experience. I wanted everyone to see the whole picture, and to see how personal everything is to me.

O conceito de *visual album* de Beyoncé foi levado ainda mais além quando lançou, em 2016, o álbum *Lemonade*. Este teve mais impacto do que outros *visual albums* anteriores (já da era digital), provavelmente, devido à sua estratégia de lançamento: o álbum foi disponibilizado apenas no *iTunes* (para *download*), no *TIDAL* (para *streaming*) — uma plataforma que requer uma subscrição paga para oferecer acesso aos seus conteúdos — e em formatos físicos, o que fez com que fosse mesmo necessário pagar para ouvir todas as canções. Foi também emitido na *HBO* apenas no dia de lançamento, conseguindo uma média de 787 mil espectadores²⁹. Trata-se de um filme de 1 hora e 5 minutos, com uma narrativa estruturada sobre o processo por que se passa para superar uma traição numa relação amorosa. Entre cada canção há imagens acompanhadas de uma narração que oferece contexto ou uma ligação para a canção seguinte. É bastante evidente a sequência de emoções durante o álbum — indiferença, ódio, raiva, tristeza, perdão, esperança e alegria —, não só por causa da ordem pensada para as faixas (que já antes se fazia em álbuns conceptuais), mas, principalmente, por causa das imagens e do contexto imersivo e narrativo em que as faixas se encontram.

Apresentarei aqui a minha experiência com este trabalho, pois considero-a relevante para explicar o impacto deste método de criação e lançamento de álbuns na indústria musical dos dias de hoje. Quando ouvi algumas canções isoladas antes de assistir ao álbum completo, não obtive grande empatia com nenhuma delas e não consegui, sequer, focar a atenção necessária para construir uma interpretação pessoal para a sua mensagem (um dos factores que terá contribuído para esta reacção é, eventualmente, o facto de eu não ser um seguidor habitual do trabalho de Beyoncé). No entanto, quando assisti ao álbum completo, as canções passaram a fazer sentido para mim, pois encontrei-me imerso numa narrativa que as unia a todas, atribuindo-lhes um contexto que é fundamental para entender a verdadeira mensagem da música. Pode comparar-se esta experiência à da visualização de um filme: ao assistir-se a uma cena isolada do filme, é possível

29 Variety. <https://variety.com/2016/tv/news/game-of-thrones-premiere-ratings-season-6-1201759863/> (data de consulta: 10.06.2019)

atribuir um significado completamente diferente do original; no entanto, ao ver-se o filme completo, a cena insere-se no contexto geral do filme e o seu verdadeiro significado torna-se perceptível — Alfred Hitchcock explica este fenómeno, denominado de efeito *Kuleshov*, no episódio *A Talk with Hitchcock* da série *Telescopes*, produzida em 1964 pela *CBC TV* (Fig. 9), ao falar da terceira das suas três principais técnicas sobre edição de filmes³⁰.

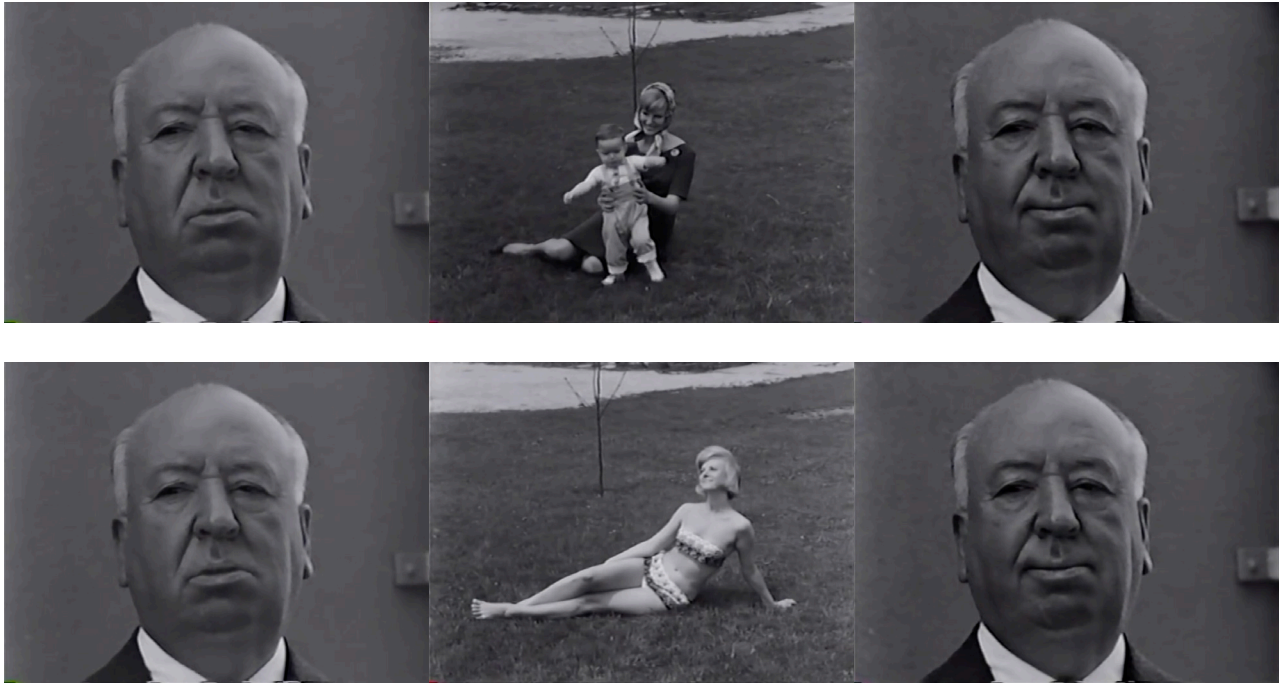


Figura 9 — Screenshots do vídeo citado. Em cima, Hitchcock é interpretado como um homem gentil e em baixo como um homem perverso, apesar de as duas cenas serem iguais e só mudar a do meio.

Este conceito, quando transposto para o contexto musical, é particularmente interessante para renovar o conceito de álbum como um todo, nomeadamente o de álbum conceptual, numa era dominada por *playlists* e na qual as pessoas raramente ouvem álbuns completos, mas sim canções retiradas do seu contexto.

2.4 Contribuição das editoras discográficas no papel do design na música

Outro aspecto que se revela interessante abordar no contexto desta dissertação é a do contributo das editoras discográficas no papel do design na música, uma vez que se pretende abordar a possibilidade de alargar o sistema criado neste projecto ao conceito de uma editora de música, em que os artigos produzidos pelos artistas estão relacionados com esse conceito.

³⁰ YouTube. <https://www.youtube.com/watch?v=MJQE7Kv-9JU> (data de consulta: 11.01.2018)

Em 1939, surgiu a editora discográfica de jazz *Blue Note Records*, fundada em New York por Alfred Lion e Max Margulis. Entre os anos 1950 e inícios de 1960, Reid Miles desenhou e fotografou quase 500 capas de discos da *Blue Note* (Bicker, 2014, p. 71), atribuindo-lhe uma identidade coesa que se tornou mais famosa do que a própria música (Fig. 10), criando um estilo internacional que ainda hoje serve de referência³¹. Reid Miles preferia a música clássica em vez de jazz, então baseava-se apenas nas indicações de Alfred Lion sobre o tom e as intenções de cada álbum para criar as capas e não na música em si (Bicker, 2014, p. 72). No artigo *Cool, clear, collected* presente na *Eye Magazine* do Outono de 1990, Robin Kinross refere que o facto de Miles não estar totalmente envolvido no conteúdo do disco permite explicar o esplêndido sentido de certeza que muitas das suas capas apresentam, pois vai directamente à solução simples e corajosa, sem ter de se confrontar com problemas que trazem demasiadas possibilidades consigo³².



Figura 10 — Capas de Reid Miles. Da esquerda para a direita, de cima para baixo: *Speakin' My Piece* de Horace Parlan, *In 'n Out* de Joe Henderson, *Unity* de Larry Young, *Let Freedom Ring* de Jackie McLean, *Hub-Tones* de Freddie Hubbard e *Happy Frame Of Mind* de Horace Parlan.

Em 1960, Creed Taylor lançou a editora *Impulse! Records*, também dedicada ao jazz, mas numa perspectiva mais ousada e experimental, tanto na música que editava como a nível visual (Bicker, 2014, p. 74). John Coltrane aproveitou a postura da editora para gravar 20 álbuns com total liberdade criativa. Os álbuns desenhados por Robert Flynn promoviam

31 Eye Magazine. <http://eyemagazine.com/feature/article/think-of-your-ears-as-eyes> (data de consulta: 16.06.2019)

32 Eye Magazine. <http://www.eyemagazine.com/%20feature/article/cool-clear-collected> (data de consulta: 16.06.2019)

o artista e veiculavam a sua mensagem, sugerindo sempre a marca da *Impulse!* (Fig. 11). As capas de Flynn, juntamente com as capas desenhadas por Reid Miles, contribuíram para o estabelecimento de uma representação gráfica do jazz rapidamente reconhecida ainda hoje (*Ibidem*, p. 75).



Figura 11 — Capas de Robert Flynn. Da esquerda para a direita, de cima para baixo: *Africa/Brass* de John Coltrane, *The Blues And The Abstract Truth* de Oliver Nelson, *Percussion Bitter Sweet* de Max Roach, *Out Of The Cool* de The Gil Evans Orchestra, *Jazz Goes To The Movies* de Manny Albam e *The John Coltrane Quartet Plays* de John Coltrane.

A *ECM Records* (*Edition of Contemporary Music*) foi fundada em 1969 por Manfred Eicher. Segundo Bicker (*Ibidem*, p. 77):

No universo actual da música, mais do que nos dos livros, é difícil reconhecer uma etiqueta pelas suas capas. Os casos como o de Reid Miles que, como vimos, dotou a *Blue Note* de uma identidade coesa e confiante, que se tornou mais famosa do que a música, criando um estilo internacional, não são comuns. (...) Esta etiqueta independente [ECM] mantém, há 45 anos, sob a direcção do seu criador, uma consistência difícil de igualar na produção de design para a indústria da música.

Durante 25 anos foi Barbara Wojirsch a principal responsável pelo design da *ECM* e foi quem definiu o estilo visual da editora até hoje presente, em que as capas vivem da combinação de fotografias, que sugerem mais do que apresentam, com tipografia neutra e minimalista (Fig. 12). Mais tarde, em 1978, entrou na *ECM* o designer Dieter Rehm, que trouxe consigo uma mudança na utilização da fotografia nas capas: até aos anos 1990, as fotografias presentes nas capas eram maioritariamente narrativas

e representacionais, chegando mesmo a ilustrar o título do álbum; depois tornaram-se “objectos de arte fotográfica que revelavam o seu significado apenas depois de uma inspecção mais próxima, atraindo o observador para um enigmático labirinto de interpretações” (*Ibidem*, p. 80).

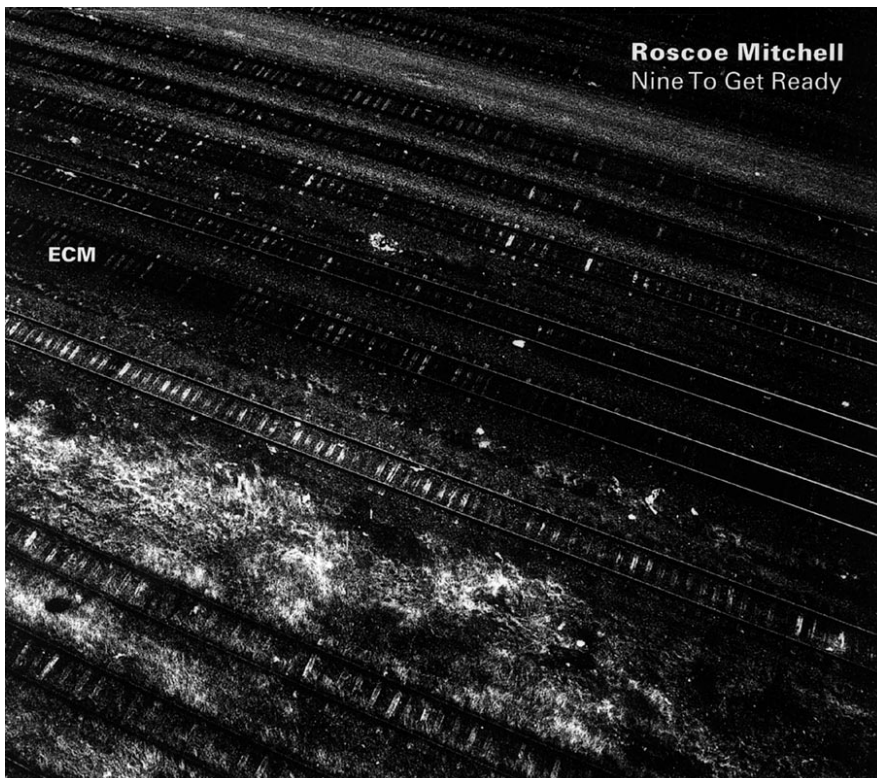
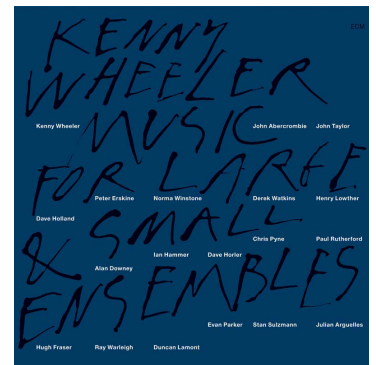
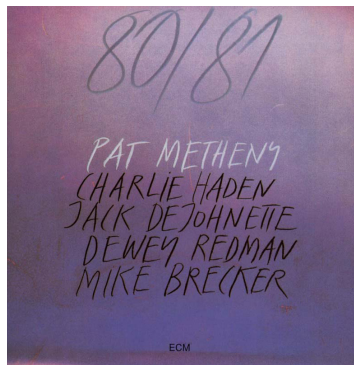
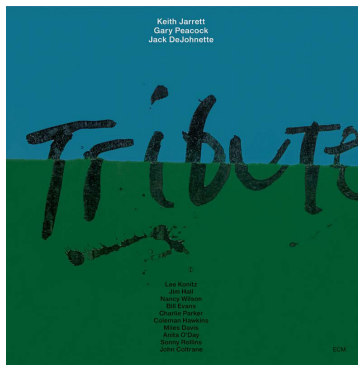
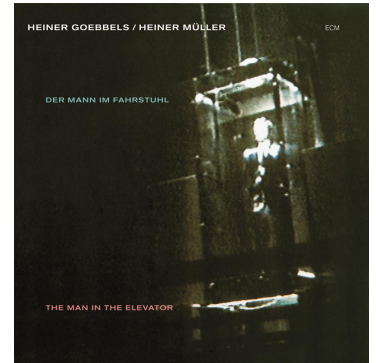
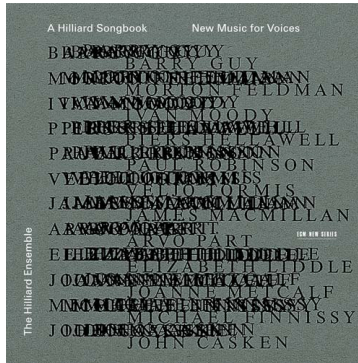


Figura 12 — Capas de Barbara Wojirsch e Dieter Rehm. Da esquerda para a direita, de cima para baixo: *A Hilliard Songbook* — *New Music for Voices* de The Hilliard Ensemble (capa: Wojirsch), *Snow* de Stephan Micus (capa: Rehm), *The Man In The Elevator* de Heiner Goebbels e Heiner Müller (capa: Rehm), *Tribute* de Keith Jarrett (capa: Wojirsch), *80/81* de Pat Metheny, Charlie Haden, Jack DeJohnette e Mike Brecker (capa: Wojirsch), *Music For Large & Small Ensembles* de Kenny Wheeler (capa: Wojirsch) e *Nine To Get Ready* de Roscoe Mitchell (capa: Rehm).

A *Factory Records* foi fundada em 1978 por Tony Wilson e Alan Erasmus e o design era da responsabilidade de Peter Saville. O trabalho de Saville foge ao paradigma colaborativo ou de interpretação presente nas editoras mencionadas nesta secção. As suas capas representavam a interpretação que Saville fazia do tempo em que os discos eram produzidos e não da sua música (Fig. 13). Os músicos não tinham opinião sobre as capas e, muitas vezes, só viam o resultado final quando já estavam na loja (*Ibidem*, p. 89).



Figura 13 — Capas de Peter Saville. À esquerda *Movement* e à direita *Procession / Everything's Gone Green*, ambos dos New Order,

A *4AD* é uma editora independente fundada em 1979 por Ivo Watts-Russell e a ela está associado o nome do designer Vaughan Oliver. A abordagem de Oliver (Fig. 14) pode ser considerada oposta à de Reid Miles e de Peter Saville, no sentido em que, para ele, cada projecto de design começa sempre pela leitura das letras e é conduzido pela música, pois, se não se liga com a música, não faz sentido (*Ibidem*, p. 85). O próprio Oliver refere, numa entrevista ao *website* cultural *The Quietus*, que o trabalho de Saville na *Factory Records* era uma antítese do seu, pois as suas capas não têm nada a ver com a música, mas sim consigo próprio e com a cultura popular do momento. Na sua opinião, qualquer *packaging* deve relacionar-se com o seu conteúdo, pois só assim é que tem substância e garante longevidade³³. Para ele, as capas devem ser uma porta de entrada para aquilo que a música é, sem a tentar definir, mas criando uma disposição e atmosfera sugestivas (Bicker, 2014, p. 85).

Estas são diferentes abordagens que vale a pena conhecer, para melhor enquadrar a via que se pretende para o presente projeto. No sistema criado, adopta-se uma abordagem oposta à de Saville, uma vez que este nem ouve a música quando cria as capas relacionadas. Também não se segue a abordagem gráfica dos designers Barbara Wojirsch e Dieter Rehm, na medida em que o grau de abstração que impõem aos seus produtos tornam-nos em fontes de interpretação quase independentes da música que os inspirou.

³³ The Quietus. <https://thequietus.com/articles/03454-vaughan-oliver-a-portrait-of-the-artist-as-a-jung-man> (data de consulta: 17.06.2019)

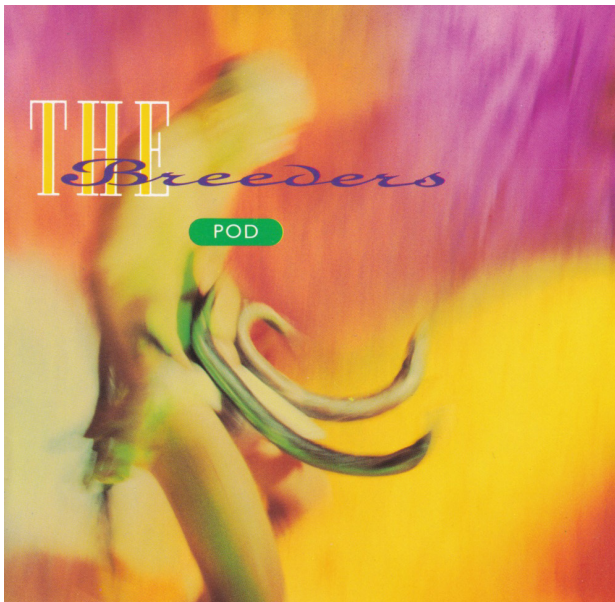


Figura 14 — Capas de Vaughan Oliver. Da esquerda para a direita, de cima para baixo: *Down Colorful Hill* dos Red House Painters, *Livonia* dos His Name Is Alive, *Pod* dos The Breeders e *Come To Pilgrim* dos Pixies.

No entanto, estes designers serviram de inspiração para este projeto, na medida em que se distinguiram pelo facto de terem gerado uma identidade na *ECM Records* que era facilmente reconhecível em qualquer dos artigos que produzia. O sistema desenvolvido neste projecto permite gerar uma identidade para uma editora recorrendo a um mesmo vocabulário visual imposto a todos os artigos produzidos. Em termos de imagem gráfica, é um projecto que está mais de acordo com a perspectiva de Vaughan Oliver na *4AD*, pois o grafismo relaciona-se directamente com a música, seja por meio de imagens concretas relacionadas com a letra da música, seja pela ambiência visual criada relacionada com a sonoridade.

2.5 Mudança de paradigma na relação concertos-álbuns de música

Antes da possibilidade de gravação de música a partir da já anteriormente mencionada invenção do fonógrafo por Thomas Edison, os artistas eram remunerados exclusivamente pelas suas performances ao vivo. Quando se tornou possível gravar música, a indústria discográfica foi crescendo de tal forma que ultrapassou consideravelmente a indústria dos concertos a nível de rendimento (El Gamal, 2012, p. 6). No entanto, as duas indústrias sempre funcionaram em paralelo, na medida em que os concertos serviam o principal propósito de promover os álbuns lançados pelo artista. Com o aparecimento da Internet, no entanto, este paradigma inverteu-se e os álbuns de música são já pensados em vários níveis para promover os concertos. Um perfeito exemplo deste facto, é o álbum *Colors* de Beck, que é constituído apenas por músicas *pop*, criadas especificamente para serem tocadas ao vivo, tal como o designer Jimmy Turrell revelou numa entrevista à *Deezer*³⁴.

A inversão da tendência das vendas nas indústrias fonográfica e da música ao vivo teve lugar no início dos anos 2000. Num artigo para a *ABC News*, Peter Kafka refere que as vendas na música desceram de 13 para 11.5 mil milhões de dólares entre 1999 e 2002, enquanto a indústria da música ao vivo subiu de 1.3 para 2.1 mil milhões de dólares entre 1998 e 2002³⁵. Esta inversão fez com que o principal rendimento dos artistas passasse a vir dos concertos (El Gamal, 2012, p. 61). O crescimento da Internet possibilitou que os artistas distribuíssem a sua música mais facilmente e, desta forma, construíssem a sua própria *fan base*. Assim, tornou-se possível adoptar novas estratégias com vista na promoção dos seus concertos. Uma delas consiste na disponibilização gratuita das músicas na Internet, aliada à utilização eficiente das redes sociais (*Ibidem*, p. 62).

El Gamal cita um artigo de 2012, escrito por Dewenter, Haucap e Wenzel (Dewenter *et al.*, 2012, p. 3), quando estabelece uma analogia entre a relação que existe entre os álbuns de música e os concertos associados e a relação que existe entre os livros e os filmes baseados nesses livros. A leitura de um livro — neste contexto, análogo à audição de um álbum de música — poderá elevar o interesse na visualização do filme relacionado — no caso, os concertos relativos ao álbum de música —, enquanto o filme irá, provavelmente, aumentar o número de cópias vendidas do livro. Tendo esta relação por base, defende-se que, apesar de a indústria da música ao vivo estar a render mais na era digital do que a fonográfica, ambas devem estar integradas num só modelo de negócio, de modo a que ambas sejam mais lucrativas (El Gamal, 2012, p. 38-39).

34 Deezer. <https://www.deezer.com/2017/09/30/beck-album-artwork-colours-customisable-design/> (data de consulta: 22.01.2018)

35 ABC News. <http://abcnews.go.com/Business/story?id=86535&page=1> (data de consulta: 22.01.2018)

2.6 Festivais vs. Concertos individuais

Um fenómeno resultante deste crescimento da indústria da música ao vivo foi a elevada ascensão dos festivais, superando até os concertos individuais. Uma das possíveis explicações para este fenómeno consiste na diferença que existe entre a relação quantidade-preço de um concerto individual e a de um festival: um consumidor poderá preferir gastar, por exemplo, 300 dólares por um festival com cerca de 100 bandas e artistas do que 90 dólares por um concerto com apenas uma banda/artista (*Ibidem*, p. 62-63).

No entanto, devem ser considerados alguns aspectos que revelam uma grande diferença na experiência de frequentar um festival ou um concerto individual. O público que assiste a um artista varia muito nos contextos de festival e de concerto próprio. Por exemplo, é mais provável num contexto de concerto individual de um artista do que num festival que o público conheça o seu trabalho e que queira adquirir o seu álbum ou *merchandise*. Além disso, o tempo de duração de uma actuação de um artista num festival é bastante mais limitado do que num concerto próprio.

Contudo, uma das principais diferenças que existe entre os dois tipos de eventos, e provavelmente o mais pertinente no contexto desta dissertação, é o facto de o foco central de um festival ser a identidade do próprio festival ao invés da dos artistas. O conceito de festival consiste numa experiência de vários dias integrada numa cultura própria (*Ibidem*, p. 63) à volta da qual se define toda a identidade. Num concerto individual, a identidade do artista é que está presente em todos os aspectos do evento, pelo que poderá desenvolver de forma mais conceptual todos os pormenores que se inserem na produção do seu espectáculo.

2.7 Breve história dos efeitos visuais em concertos de música

Desde a revolução do *rock and roll*, que teve início nos anos 1950, que a identidade dos artistas passou a ter mais relevância em toda a estratégia de design e de marketing do seu trabalho. Nos finais da década de 1960, essa identidade passou a ser tida em conta também na concepção dos espectáculos ao vivo, mais propriamente na cenografia e na elaboração dos efeitos visuais.

No início, esta vertente consistia apenas em efeitos rudimentares, como, por exemplo, os utilizados nos concertos dessa altura de Alice Cooper. Pouco mais tarde, já na década de 1970, bandas como Led Zeppelin, Pink Floyd e Genesis foram pioneiras no tipo de efeitos visuais ainda hoje utilizados, embora de forma consideravelmente mais desenvolvida. Os Led Zeppelin foram a primeira banda a utilizar um laser como efeito visual, em 1975 (Fig. 15), ideia esta adaptada posteriormente pelos The Who, ao manusear um laser que atravessava uma rede de difracção, dividindo-o em vários.

Os efeitos visuais dos concertos dos anos 1970 e 1980 foram dominados por explosões e acrobacias, exploradas por bandas e artistas como o Michael Jackson, Pink Floyd, Kiss, AC/DC e Queen³⁶.

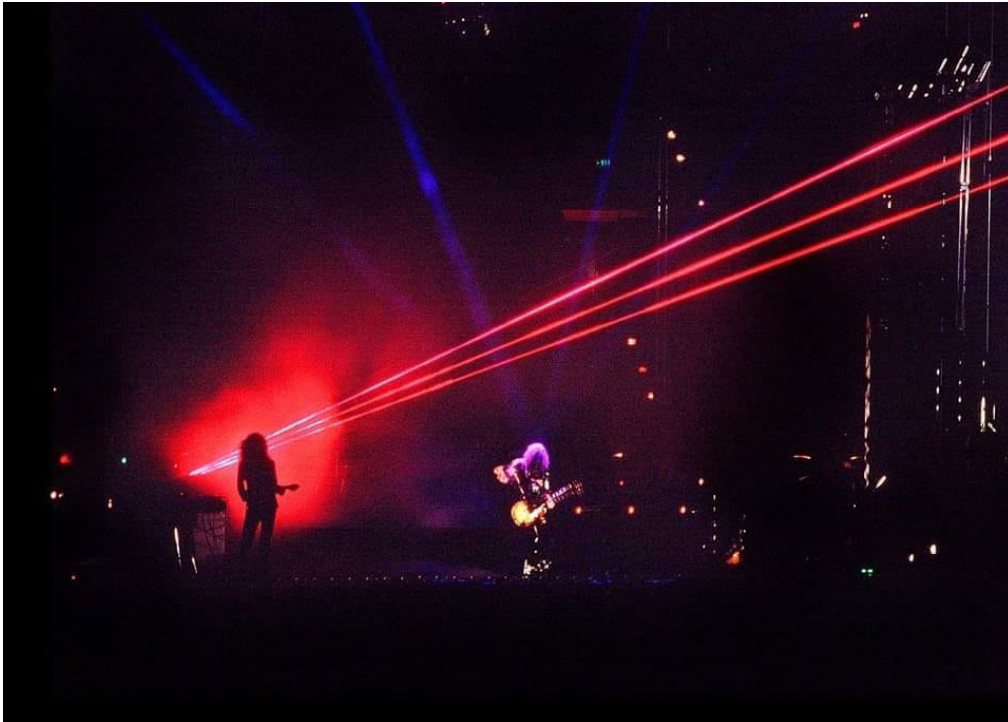


Figura 15 — Efeitos visuais de um concerto dos Led Zeppelin em Seattle, em 1975.

No início dos anos 1990, já em plena revolução digital, os U2 elevaram os efeitos visuais com a utilização de ecrãs gigantes formados por diversas televisões dispostas em cima umas das outras. Também Madonna começou a utilizar vídeo nos seus concertos, para sustentar as histórias que contava em palco. As tecnologias digitais e robóticas já dominavam os efeitos visuais dos concertos dos anos 2000, quando, por exemplo, a banda Daft Punk elevou a utilização de lasers dos Led Zeppelin e The Who a um nível mais sofisticado, pois criavam formas concretas, como pirâmides. O crescimento dos efeitos visuais nos concertos tem vindo a ser exponencial com o desenvolvimento das novas tecnologias, como se poderá observar em concertos de artistas como U2, Beyoncé, Coldplay, etc.³⁷.

2.8 Novos métodos de actuação e de interacção com o público

Todas estas inovações fizeram com que houvesse uma ascensão do conceito de concerto desde a performance de músicas para todo um espectáculo

³⁶ The Washington Post. https://www.washingtonpost.com/entertainment/music/how-concerts-shifted-from-songs-to-spectacles/2014/05/22/ca521340-d6ce-11e3-8a78-8fe50322a72c_story.html (data de consulta: 22.01.2018)

³⁷ *Idem*.

teatral e audiovisual. No entanto, a inovação na concepção dos concertos está longe de estabilizar, havendo já diversos artistas a trabalhar com novas tecnologias a fim de criar novos métodos de actuação. Alguns exemplos são **i)** o designer de som Ross Flight, que utiliza o sensor de movimento *Kinect* da *Microsoft* para criar música a partir do movimento do corpo; **ii)** a coreógrafa Laura Kriefman, que cria música através de dança; **iii)** Tim Exile, um produtor que inventou os seus próprios instrumentos electrónico; **iv)** ou Beardyman, um *beatboxer* que construiu uma máquina de produção de música em *real-time* apenas com sons da voz. Imogen Heap apresentou em 2011 um par de luvas (Fig. 16) que lhe permitia gravar, ampliar ou fazer *loops* de instrumentos acústicos ou da sua voz, ou ainda tocar instrumentos virtuais, manipulando todos estes sons ao vivo³⁸.



Figura 16 — Imogen Heap a demonstrar a o funcionamento das Mi.Mu Gloves na *WIRED* 2012.

Além das inovações tecnológicas que se têm desenvolvido para criar diferentes formas de actuação, também têm surgido as destinadas à interação do artista com o seu público. Imogen Heap utilizou em 2009 o auditório online *Vokle* para fazer audições aos seus seguidores a partir de vídeo, a fim de escolher um para tocar consigo numa *tour*. Têm sido criadas também algumas ferramentas como os *smart earbuds*, desenvolvidos pelos Doppler Labs, que permitem personalizar a experiência de audição num concerto a partir do ajuste do volume e equalização ou a aplicação de efeitos no som; ou a pulseira *The Basslet*, que permite os ouvintes sentir no corpo a música que ouvem nos seus dispositivos de música a partir de estímulos, como se estivessem a ouvi-la no contexto de um concerto ³⁹.

³⁸ NEXT. <https://howwegettonext.com/the-history-and-future-of-live-music-147ecde437b7> (data de consulta: 23.01.2018)

³⁹ *Idem*.

Também as redes sociais têm fornecido ferramentas que permitem a assistência aos concertos a partir de todo o mundo através de *live streaming*. Além disso, permitem a partilha da experiência individual das pessoas presentes a partir de ferramentas como as *stories* do *Snapchat* ou do *Instagram*. Aliada a estas inovações das redes sociais, estão as tecnologias de realidade virtual, que elevam a experiência de assistência dos concertos em casa a um nível bem mais realista⁴⁰.

Todas estas tecnologias levam a crer que os paradigmas na indústria da música ao vivo irão sofrer mudanças significativas, pelo que os artistas terão de investir em novas formas de apelar à presença do seu público nos concertos em vez de experienciar em casa. A inovação nos efeitos visuais aliados à música poderá ser uma estratégia válida, pois o ambiente audiovisual imersivo presente num concerto não se experiencia da mesma forma no local do concerto ou em casa, a partir de instrumentos electrónicos.

2.9 Conclusões

Neste capítulo estudou-se a história das indústrias fonográfica e da música ao vivo com o objectivo de conhecer as estratégias de design e marketing abordadas em cada momento e em cada uma em função das novas tecnologias que foram surgindo ao longo do tempo. Como foi possível referir, com a evolução tecnológica, foram acontecendo verdadeiras mudanças de paradigma na indústria da música e da sua relação com o design e as novas tecnologias.

Neste estudo, a história desta indústria e a sua relação com o design e a tecnologia permite uma apropriação do contexto em que se insere o problema para o qual se procurou dar resposta com a componente prática deste projecto. É particularmente importante ter em consideração a rapidez com que as últimas alterações têm acontecido, sendo certo que as formas de divulgação da música de cada músico, bem como da sua mensagem, não têm acompanhado a mesma rapidez, havendo dificuldade em chegar a um público habituado a *playlists* com músicas de diversos músicos, verificando-se uma verdadeira desintegração dos álbuns produzidos por cada músico. São ainda referidas neste capítulo algumas mudanças de paradigma na relação entre os álbuns gravados e os respectivos concertos, defendendo-se actualmente que devem ser pensados de modo integrado. Refere-se a grande adesão do público de hoje aos festivais de música em detrimento dos concertos individuais, o que constitui um desafio para cada músico encontrar uma forma de chegar ao seu público. Alguns casos na história recente dos concertos individuais em que alguns músicos têm investido nos efeitos visuais que integram no espectáculo ao vivo têm oferecido sucesso que foi considerado na ideia do presente projecto. Regista-se também neste capítulo algumas notas sobre as novas formas de interacção do músico com o seu público, com recurso às novas tecnologias (por exemplo da realidade virtual) e às redes sociais que trazem novas formas de experienciar a música ao vivo às quais os músicos não podem ficar indiferentes.

Casos de Estudo

Capítulo III

- 3.1 Geração de imagens a partir de texto
- 3.2 Geração de som a partir de texto
- 3.3 Ilustração generativa
- 3.4 Modelação de imagem de acordo com som
- 3.5 Ferramentas de geração de capas
- 3.6 Álbuns de música
- 3.7 Conclusões

Capítulo III

Casos de Estudo

A pesquisa de projectos relacionados com a dissertação permite conhecer um pouco do que tem sido feito na área de estudo, analisando pontos fortes e fragilidades de cada um. São apresentadas algumas abordagens interessantes que podem constituir-se como fontes inspiradoras e outras não tão bem sucedidas que devem ser evitadas. Em qualquer dos casos, são referências para as diversas matérias abordadas na proposta de resolução do problema que se aborda neste projecto.

Começa-se por descrever trabalhos pertencentes a outras áreas cujas capacidades, se aliadas à indústria da música, poderão vir a acrescentar características que podem vir a ajudar os artistas a apelar a atenção do público. Posteriormente, referem-se exemplos de artistas de música que têm já estes problemas em conta e desenvolveram possíveis soluções para esta problemática, relacionadas com as que se pretendem encontrar durante o desenvolvimento deste trabalho.

Para garantir eficiência na análise dos casos de estudo seleccionados, procedeu-se a uma organização por diferentes secções correspondentes a temáticas relacionadas com o projecto: **i)** geração de imagens a partir de texto, **ii)** geração de som a partir de texto, **iii)** ilustração generativa, **iv)** modelação de imagem de acordo com som, **v)** ferramentas de geração de capas e **vi)** alguns exemplos de imagens gráficas de álbuns de música e, em alguns casos, dos efeitos visuais dos concertos correspondentes. Começa-se por abordar projectos mais técnicos do que gráficos, passando, posteriormente, por projectos que misturam a componente técnica com a gráfica e termina-se com projectos puramente gráficos. Esta é também a ordem que se seguiu no desenvolvimento da dissertação: da componente mais técnica para a componente mais gráfica.

3.1 Geração de imagens a partir de texto

Os projectos descritos nesta secção revelam-se importantes para conhecer abordagens já testadas na obtenção de imagens a partir de texto, uma das etapas da componente prática da presente dissertação. O primeiro projecto corresponde a um algoritmo desenvolvido para encontrar as imagens mais adequadas para um determinado bloco de texto dentro de um estilo gráfico seleccionado. O segundo consiste na geração de símbolos em *real-time* relativos a uma história narrada frase-a-frase.

Auto-Illustrating Poems and Songs with Style

Este é um projecto de investigação desenvolvido por Katharina Schwarz e Hendrik P. A. Lensch da University of Tübingen e por Tamara L. Berg da University of North Carolina. Consiste numa *framework* que, tal como o título indica, permite a ilustração automática de poemas e canções com um estilo visual especificado pelo utilizador, mantendo sempre relevância semântica e coerência visual (Schwarz *et al.*, 2017). Para a demonstração do método utilizado no projecto, Schwarz recolheu 200 poemas e canções da Internet e trabalhou à volta de 14 milhões de imagens do Flickr (*Idem*). A partir de um complexo algoritmo, é possível criar uma sequência de imagens, todas com um estilo idêntico, em que cada uma representa visualmente o que está escrito em cada verso (Fig. 17).

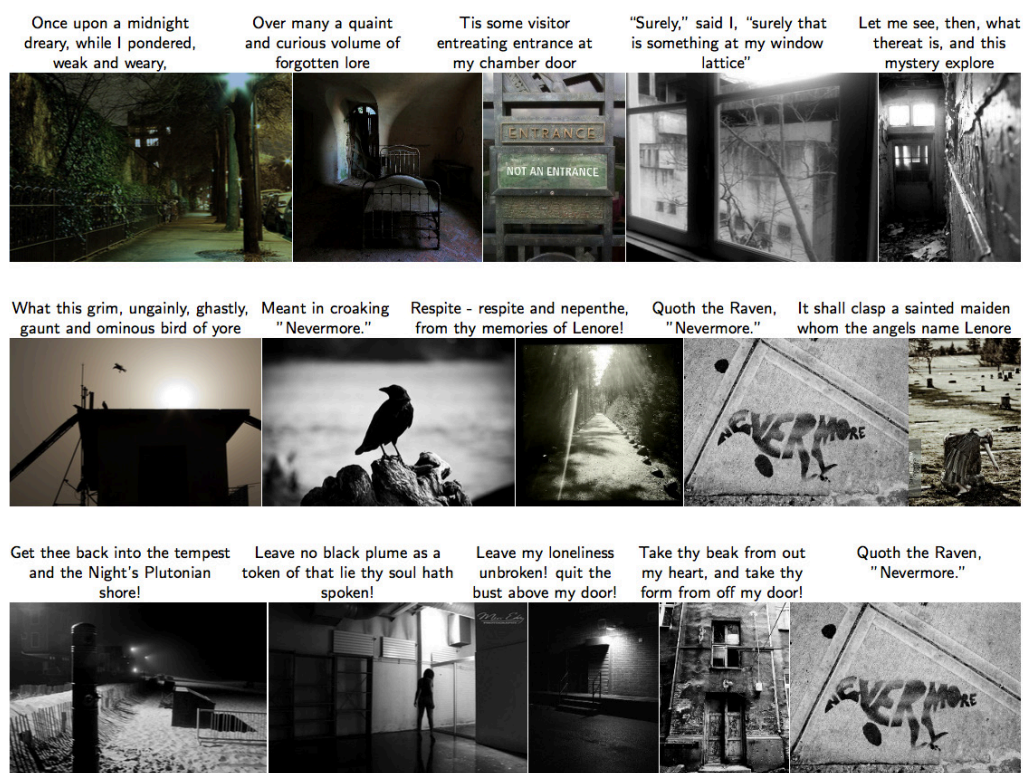


Figura 17 — Ilustração do poema *The Raven* de Edgar Allen Poe no estilo seleccionado "noir". Screenshot do artigo.

Este projecto é particularmente relevante para a componente prática desta dissertação, tendo em conta que ambos têm por base a ilustração automática de poemas ou canções. Uma das virtudes deste projecto é a ilustração da ideia geral de cada verso, em vez de cada palavra, mantendo uma sequência representativa do poema completo. Este é um aspecto pertinente que foi considerado na presente dissertação, devido ao elevado tempo de processamento que a ilustração de cada palavra exige. No entanto, o funcionamento do resultado final como vídeo não é aprofundado, levando a crer que se trata apenas de uma sequência de imagens, tornando o projecto menos interessante a nível de experiência visual.

*The Shape of Story: A Semiotic Artistic Visualization
of a Communal Storytelling Experience*

Este projecto, desenvolvido por Duri Long e Sanjana Gupta no Georgia Institute of Technology, consiste num sistema capaz de entender discurso falado em inglês e representá-lo em *real-time* através de uma linguagem visual baseada em símbolos (semiótica) (Long *et al.*, 2017). Este sistema é integrado numa instalação artística interactiva que fornece um ambiente propício para diversos participantes criarem histórias em conjunto, uma frase de cada vez (*Idem*). As histórias são representadas visualmente à medida que são narradas, o que faz com que os participantes criem, além da história, uma composição visual colaborativa (Fig. 18).

A abordagem semiótica deste projecto chegou a ser considerada para este trabalho, no entanto, a representação por símbolos torna a composição final numa narrativa abstracta, pois não permite que haja uma releitura concreta da história a partir das imagens, como o projecto descrito anteriormente — para que isso acontecesse, seria necessário recorrer a uma legenda que associasse um significado a cada símbolo — e, portanto, optou-se por descartar esta hipótese e seguir com ilustração concreta. Assim, a vertente visual deste projecto não teve relevância para o processo desta dissertação. No entanto, um dos aspectos fundamentais deste trabalho surgiu a partir da adopção da abordagem utilizada neste projecto de Duri Long no mapeamento de qualquer palavra a um determinado número fixo de símbolos — através das quatorze acções básicas definidas por Roger C. Schank (Schank, 1973) — para realizar um mapeamento dos verbos presentes na letra de uma música a um número fixo de animações.

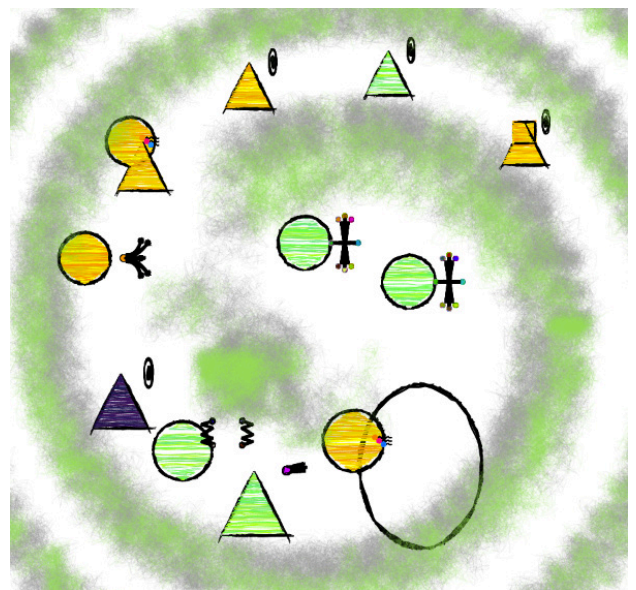


Figura 18 — À esquerda: funcionamento da experiência de *storytelling*. À direita: exemplo de *output* de uma história narrada.

3.2 Geração de som a partir de texto

Nesta secção apresenta-se um projecto, no qual se produz música tendo por base a emoção contida num bloco de texto. Este estudo revelou-se útil para a componente de modelação de imagens de acordo com som do sistema desenvolvido neste trabalho, pois revela formas de mapeamento de características do som a outras variáveis.

Sonifying Twitter's Emotions Through Music

Este projecto, realizado por Mariana Seiça na Universidade de Coimbra, tem como objectivo traduzir automaticamente emoções de utilizadores do *Twitter*, recolhidas a partir dos seus *tweets*, em composições musicais (Seiça, 2017). No artigo relativo ao projecto que foi apresentado no *CMMR 2017*, é referido que foi definido um conjunto de emoções para avaliar os *tweets* e, posteriormente, mapeou-se essas emoções a variáveis musicais, construindo, assim, composições musicais (Seiça et al., 2017). Um dos aspectos mais relevantes deste projecto para a componente prática desta dissertação é o estudo das características do som e da música e como é possível mapeá-las com outras variáveis.

3.3 Ilustração generativa

Esta secção destina-se ao destaque de um projecto que serviu de referência técnica e gráfica para o desenvolvimento da etapa da ilustração generativa presente na componente prática da dissertação.

Gold, por Memo Akten

Esta instalação de Memo Akten funciona como um “espelho mágico”, no sentido em que capta a figura do utilizador e devolve-a desenhada a partir de pequenas estrelas douradas. Quando o utilizador se mexe, a figura desenhada transforma-se numa nuvem de estrelas até estabilizar e desenhá-la a sua nova posição (Fig. 19). O projecto serve de crítica ao fenómeno “15 minutos de fama” presente na geração das redes sociais e previsto pelo Andy Warhol⁴¹. O aspecto visual deste projecto é uma referência significativa para a presente dissertação, pois funciona graficamente em estacionário e em movimento.

41 Memo Akten. <http://www.memo.tv/gold/> (data de consulta: 16.01.2018)

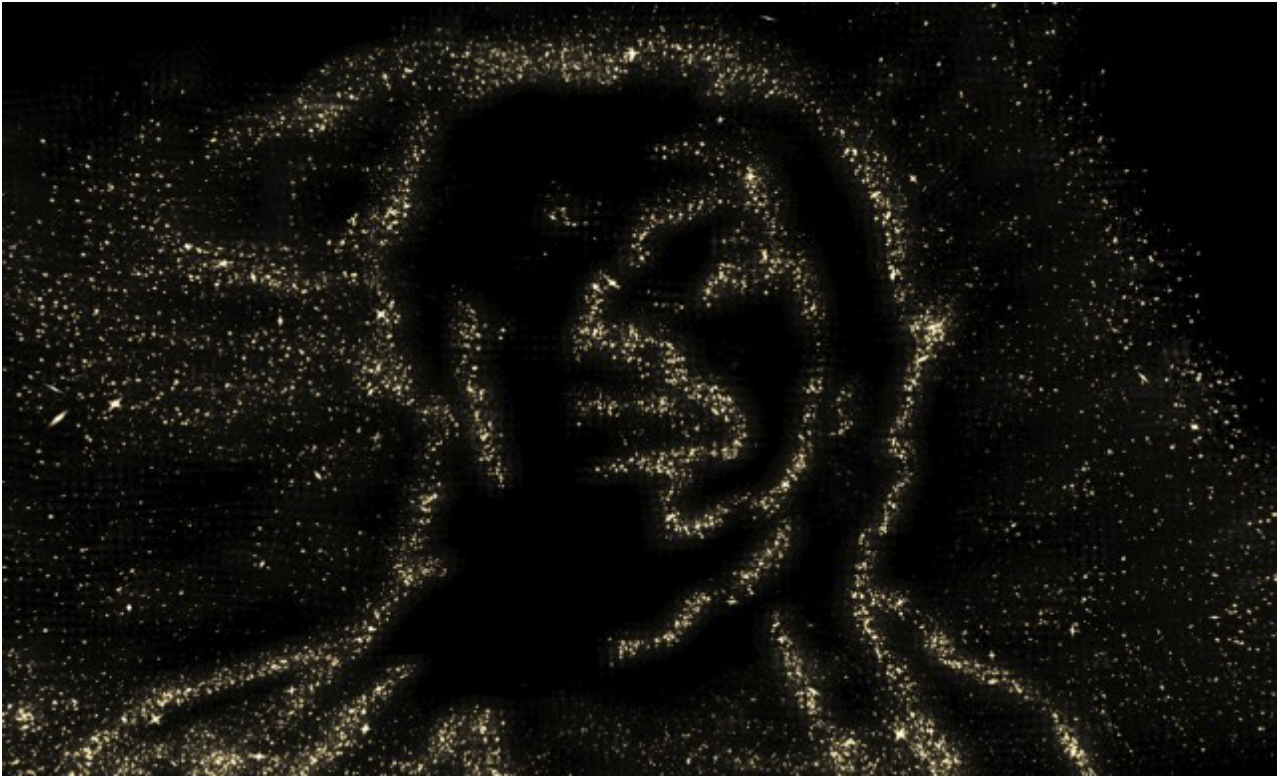


Figura 19 — Frame do vídeo de demonstração da instalação Gold.

3.4 Modelação de imagem de acordo com som

Nesta secção são descritos cinco projectos que exploram a modelação de imagem de acordo com som, também explorada na componente prática desta dissertação. Os projectos seleccionados abordam diferentes abordagens para alcançar diferentes objectivos. O primeiro revela um método que tem por base uma imagem simples e gera uma nova imagem apenas depois da modificação das variáveis sonoras (previamente mapeadas com variáveis da imagem). O segundo demonstra uma abordagem de modelação de imagem de acordo com som que funciona em *real-time*, mas sem possibilidade de interacção. O terceiro consiste numa abordagem que permite a modelação de imagem a partir da produção de som em *real-time*, o que acrescenta a componente de interacção relativamente aos dois projectos anteriores. Os dois últimos projectos apresentam exemplos concretos de projectos que utilizam modelação de imagem de acordo com som no contexto da *web*, o que abre algumas portas para uma eventual extensão deste projecto aos *browsers*. O primeiro aplica a modelação de imagem de acordo com som na tipografia, mais concretamente no que diz respeito ao conceito de *variable font*, enquanto o segundo utiliza a modelação de imagem de acordo com som para criar visualizações interactivas para todas as músicas de um álbum de música, criando um novo tipo de *visual album* interactivo.

Glitchometry, por Daniel Temkin

Nesta experiência, Temkin explora a distorção de imagens de acordo com características ou efeitos sonoros. A imagem base é sempre simples, como, por exemplo, um círculo ou um quadrado. De seguida, importa a imagem para um programa de edição de som, no qual altera variáveis, como o volume ou o *pan*, ou, ainda, aplica efeitos, como *delay* ou *flanger*⁴², a canais de cor, transformando a imagem (Fig. 20). Os resultados visuais são difíceis de prever, tendo em conta que a imagem nunca é trabalhada como imagem, mas sim como som⁴³. Este projecto é uma interessante referência para a presente dissertação, devido à manipulação de imagem através de variáveis sonoras, embora se trate de uma manipulação manual.

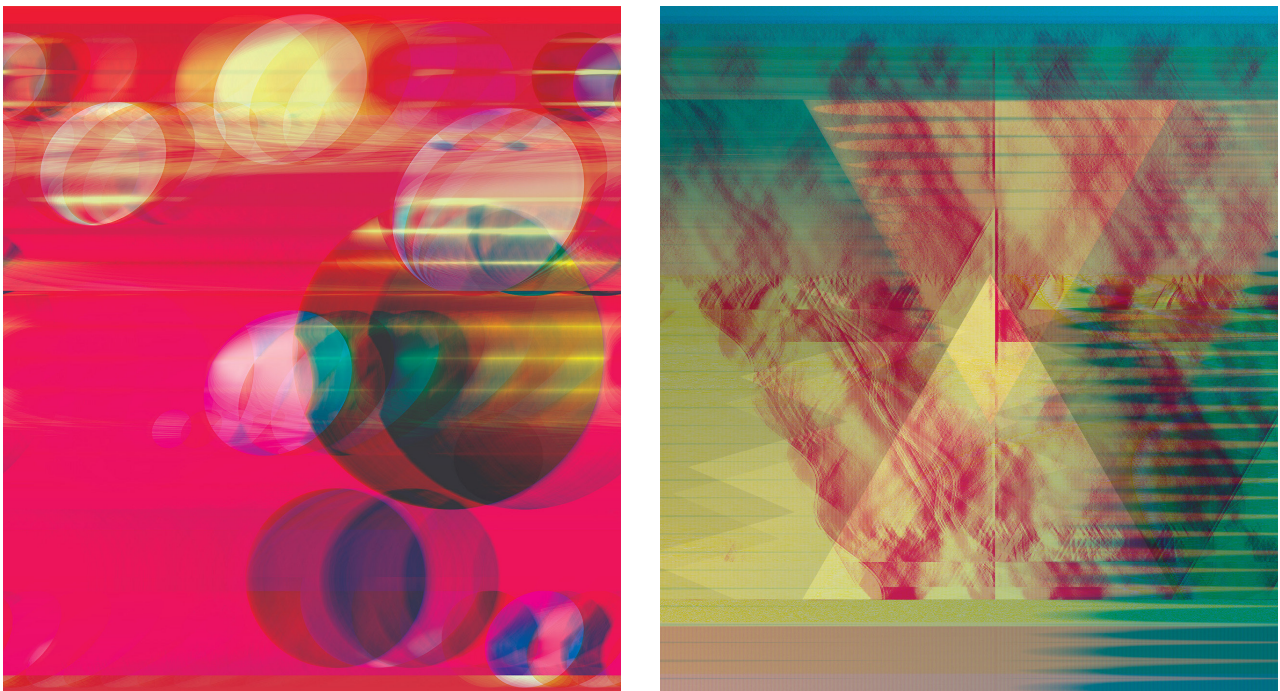


Figura 20 — Duas das imagens resultantes da experiência de Temkin, consideravelmente distintas uma da outra.

Communion, por FIELD

Communion é uma instalação artística desenvolvida pelo estúdio criativo FIELD em colaboração com Matt Pyke. A instalação tem lugar numa sala cujas paredes estão repletas de ecrãs quadrados, cada um destinado a um organismo gráfico computacionalmente gerado que reage ao som de uma banda sonora (Fig. 21). Cada criatura pequena destas tem uma forma diferente e dança com movimentos limitados pela sua anatomia. A relevância deste projecto para a presente dissertação encontra-se no mapeamento do

42 Co.Design. <https://www.fastcodesign.com/1669501/synesthesia-20-digital-images-distorted-by-sound-waves> (data de consulta: 16.01.2018)

43 Daniel Temkin. <http://danieltemkin.com/Glitchometry/Index/GlitchometryCircles4> (data de consulta: 16.01.2018)

ritmo da música ao movimento das criaturas. Quanto mais animada for a música, mais as criaturas se mexem e vice-versa, tal como se pode observar no vídeo de apresentação da instalação publicado pelo estúdio⁴⁴. Este aspecto faz com que exista uma representação visual da emoção (mais concretamente do *Arousal*) da música, tal como o que acontece na componente prática desta dissertação.

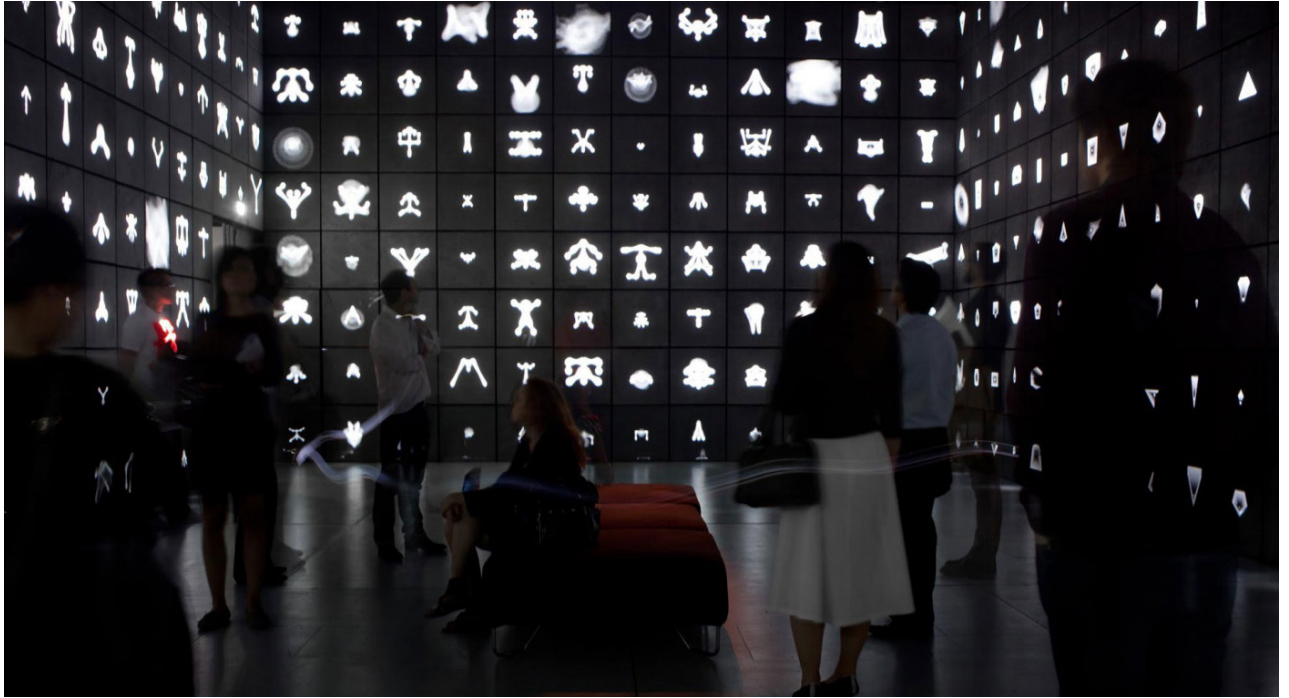


Figura 21 — Instalação *Communion* dos FIELD.

Soundfield, por The Light Surgeons

A *WIRED Magazine* encomendou uma instalação artística para o evento *Bentley Motors Night Call* no Wind Tunnels do Farnborough Airport à empresa The Light Surgeons. O resultado foi uma instalação baseada em vídeo com reacção ao som em *real-time*, utilizando a plataforma *openFrameworks*. Assim, quando as pessoas passavam pelo túnel, todos os vectores espalhados pelas paredes reagiam ao som⁴⁵ (Fig. 22). Mais uma vez, existe um mapeamento das características sonoras a variáveis visuais, desta vez com uma componente de interacção em *real-time*. Este projecto serve de referência para a componente de *real-time* que se pretende desenvolver futuramente aquando da aplicação do sistema nos efeitos visuais de um concerto. A *openFrameworks* revela-se uma ferramenta poderosa para este tipo de trabalhos, tornando-se uma potencial candidata a ser utilizada em experimentação futura deste projecto.

⁴⁴ FIELD. <https://www.field.io/project/communion/> (data de consulta: 16.01.2018)

⁴⁵ The Light Surgeons. <http://www.lightsurgeons.com/art/soundfield/> (data de consulta: 16.01.2018)

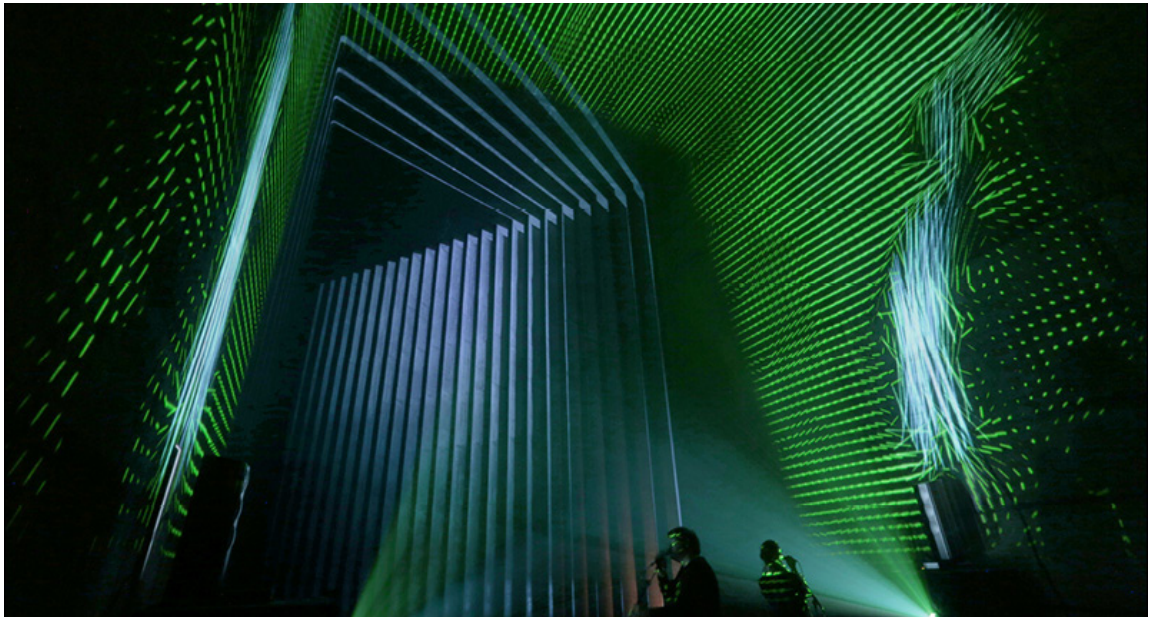


Figura 22 — Instalação *Soundfield* dos The Light Surgeons.

Phase, por Elias Hanzer

Um dos grandes fenómenos que começa a surgir, principalmente na área de *web design*, é o conceito de *variable font*. Segundo a Pangram Pangram Foundry, uma *variable font* é uma fonte que encapsula uma família inteira, ou seja, em vez de a fonte ser constituída por vários ficheiros relativos a pesos separados — por exemplo, *light*, *regular*, *bold*, etc. — a fonte contém todos os pesos no mesmo ficheiro, incluindo os intervalos entre cada peso⁴⁶, numa passagem contínua do peso mais leve para o peso mais pesado.

Levado mais longe, este conceito pode ser estendido ao próprio desenho da fonte, e não só aos pesos. A empresa suíça de design de tipos Dinamo tem estado a trabalhar neste conceito, desenvolvendo inclusive uma ferramenta em que as fontes podem ser testadas com campos de edição dos parâmetros de variação que cada fonte possibilita⁴⁷ (Fig. 23).

Elias Hanzer aplica este conceito à variação de acordo com som. *Phase* é uma tipografia generativa e variável desenhada a partir de módulos que estão na base de um infinito número de formas (Fig. 24). Pode ser manipulado em real-time a partir de controlos associados aos vários parâmetros de variação da fonte ou a partir de som⁴⁸.

Este conceito abre portas para uma possível exploração futura desta dissertação no contexto da tipografia, a fim de complementar a ilustração da letra da música com a própria letra da música de modo que ambas respondam ao som de formas distintas.

46 Pangram Pangram Foundry. <https://pangrampangram.com/blogs/journal/variable-fonts> (data de consulta: 17.06.2019)

47 Dinamo Font Gauntlet. <https://dinamodarkroom.com/gauntlet> (data de consulta: 17.06.2019)

48 Elias Hanzer. <https://www.eliashanzer.com/phase/> (data de consulta: 17.06.2019)



Figura 23 — Screenshot da ferramenta desenvolvida pelos Dinamo para testar variable fontes: *Dinamo Font Gauntlet*.

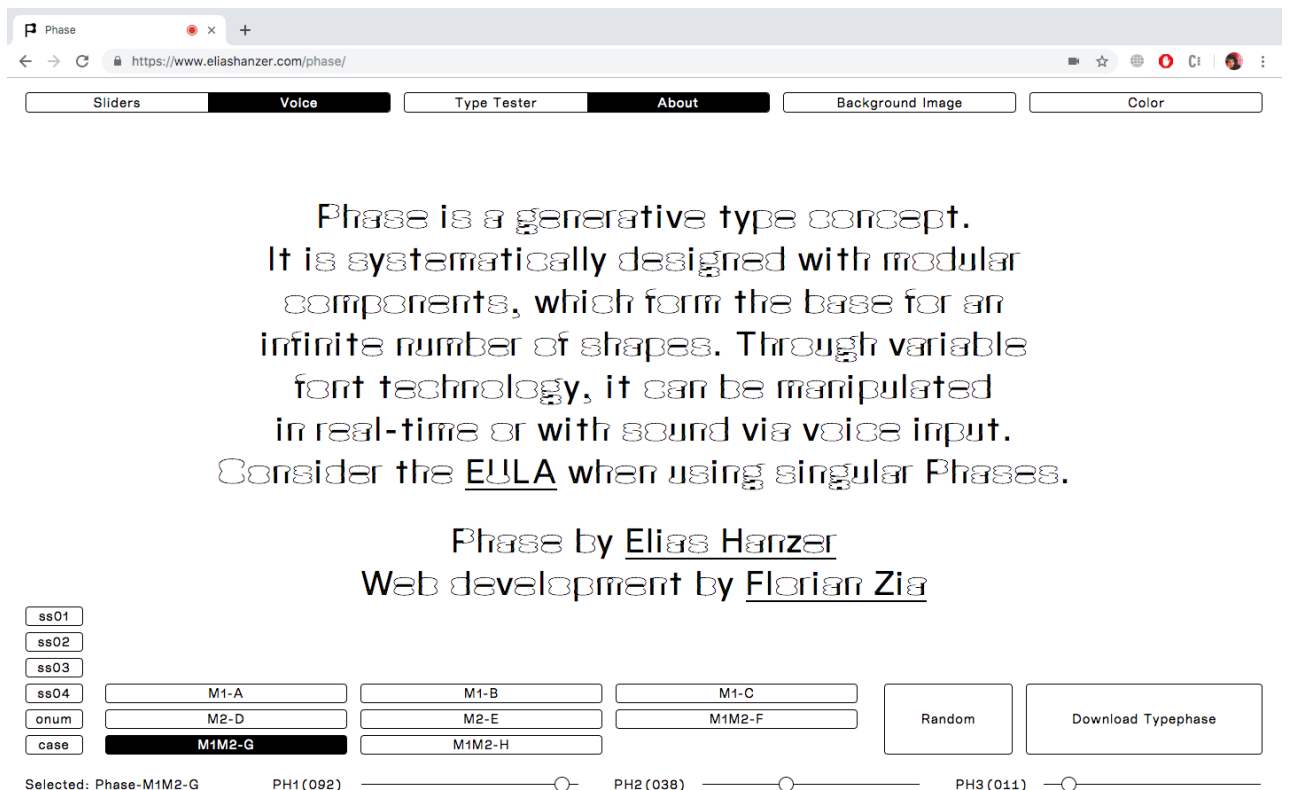


Figura 24 — Screenshot da variable font *Phaser*, criada por Elias Hanzer e que tem a possibilidade de reagir ao som.

Prologue EP, de Emmitt Fenn, por Active Theory

Emmitt Fenn lançou o seu *EP Prologue* em 2017 em formato exclusivamente online, não só nas várias plataformas online, mas também no seu *website* em forma de *visual album* interactivo⁴⁹. Quando se entra no *website*, pode-se ver a lista de músicas e, para cada uma, foi desenvolvida — pelos Active Theory — uma visualização interactiva diferente a partir da biblioteca de *JavaScript Three.js* (Fig. 25).

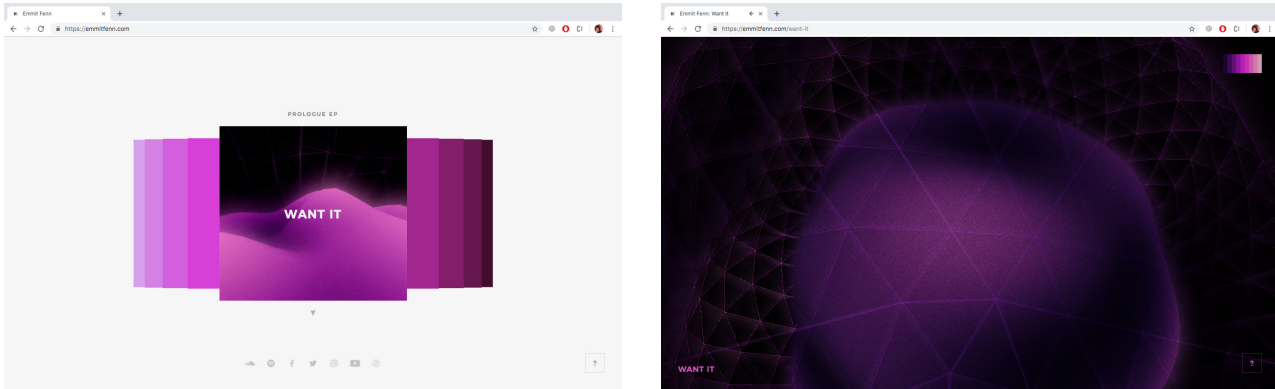


Figura 25 — Screenshots do *website* de Emmitt Fenn para o *Prologue EP*. À esquerda: lista de músicas do *EP*. À direita: *still* de uma das músicas.

Tanto este projecto como o anterior abrem portas para a exploração futura desta dissertação em ambiente de *web*, associando visualizações interactivas ao *website* do artista que estiver a aplicá-la.

3.5 Ferramentas de geração de capas

Esta secção dedica-se à apresentação de dois projectos em que se utilizaram ferramentas desenvolvidas para a geração de capas de álbuns de música em dois contextos distintos, mas nos quais o sistema criado neste trabalho pode ser aplicado. O primeiro trata-se de uma ferramenta de geração de capas criado para que uma editora discográfica criasse as capas dos discos dos seus artistas de forma autónoma, mas controlada, devido aos parâmetros definidos pelo criador da ferramenta. O segundo projecto é uma ferramenta aplicada no contexto de um álbum dos Thirty Seconds to Mars, de modo a aproximar o álbum do consumidor e torná-lo mais pessoal e, ao mesmo tempo, criar uma enorme colecção de capas espalhadas por todo o mundo.

Tonboutique Designer, por Tim Rodenbröker

A *Tonboutique Records* é uma editora discográfica baseada na Alemanha e Noruega, dedicada a música *house* e *techno*⁵⁰. Tim Rodenbröker é um designer alemão que se dedica e explora principalmente a área de *creative*

⁴⁹ Emmitt Fenn. <https://emmittfenn.com> (data de consulta: 17.06.2019)

⁵⁰ Tonboutique Records. <https://www.tonboutique-records.de> (data de consulta: 17.06.2019)

*coding*⁵¹ e desenvolveu uma ferramenta em *React.js* e *HTML5 Canvas* que permite a um utilizador — neste caso, a editora — criar capas para álbuns com base nos parâmetros de escolha presentes na ferramenta. O utilizador pode inserir o texto que deseja, seleccionar a família tipográfica, o tamanho e peso da fonte, um elemento gráfico seleccionado a partir de uma lista de elementos criados propositadamente para a ferramenta, o seu tamanho e posição e as cores do fundo e do elemento gráfico. No fim o utilizador pode visualizar o resultado num *mockup* de vinil e, por fim, exportar o resultado, obtendo um ficheiro *PNG* com alta qualidade (Fig. 26).

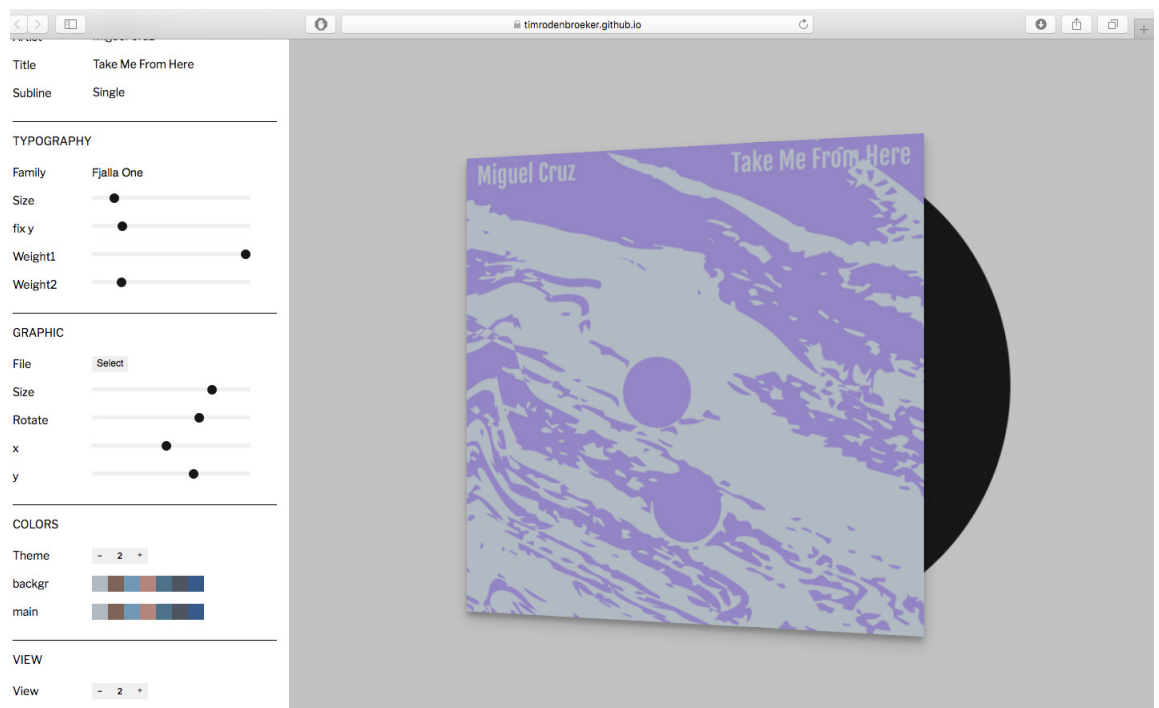


Figura 26 — Screenshot de capa exemplo criada para a canção *Take Me From Here* da minha autoria com a *Tonboutique Designer*.

Podemos associar este projecto às editoras mencionadas no capítulo anterior, principalmente a *ECM Records*, no sentido em que os artistas pertencentes à editora têm um elo de ligação entre eles que os associam à editora e à música por ela difundida. A ligação existente entre o design das capas dos discos da *ECM Records* é manual, mas, possivelmente, poderia ser obtida com este tipo de abordagem.

No contexto deste trabalho, uma editora discográfica poderia utilizar o sistema para desenvolver capas diferentes para os seus artistas, mas seguindo uma linha gráfica semelhante, tal como acontece neste projecto.

Estas ferramentas podem revelar-se bastante úteis, não só quando aplicadas no contexto das editoras discográficas, mas também para criar colecções de capas para o mesmo disco ou a mesma banda, como veremos no seguinte caso de estudo.

51 Tim Rodenbröcker. <https://timrodenbroecker.de/about/> (data de consulta: 17.06.2019)

America, dos Thirty Seconds To Mars, por Designedmemory

Em 2018, os Thirty Seconds To Mars lançaram um álbum chamado *America* cuja direcção artística foi da responsabilidade de Brian Roettinger. Acabaram por fazer uma colecção de capas puramente tipográficas, com fundos de cores diferentes, em que cada uma exibia uma lista relativa a nomes, marcas, itens americanos e conceitos derivados de inquéritos e estatísticas (Fig. 27). Por exemplo, numa das capas podia ler-se “Kylie, Elvis, Kanye, Jesus, Lebron, Oprah: Six american names” e numa outra lia-se “Apple, Google, Microsoft, Walmart, IBM: Five most valuable trademarks as reported by Forbes”. As capas todas juntas revelavam os aspectos mais relevantes da identidade dos EUA como país, segundo a banda. Jared Leto refere que as listas funcionam como cápsulas do tempo: isoladas poderão parecer provocativas, mas todas juntas conferem uma sensação da cultura da qual faz parte e dos tempos que se vivem⁵². Esta colecção poderá suscitar ao consumidor a vontade de adquirir todos os itens da colecção ou de escolher o item que contém os aspectos com que mais se identifica.



Figura 27 — Diversas versões para a capa do álbum *America* dos Thirty Seconds To Mars.

Além disso, para tornar o álbum ainda mais pessoal, a empresa Designedmemory (Matthew Daniel Siskin) desenvolveu um gerador de capas personalizadas em que o utilizador podia escrever a lista

52 It's Nice That. <https://www.itsnicethat.com/news/willo-perron-thirty-seconds-to-mars-america-album-cover-graphic-design-220318> (data de consulta: 17.06.2019)

que quisesse e escolher a sua cor preferida, terminando a lista com a frase “This is my America”⁵³ (Fig. 28).

No capítulo relativo à componente prática desta dissertação, apresenta-se uma aplicação do sistema na criação de uma colecção de capas que foi inspirada pela ideia deste projecto.

Este tipo de ferramentas de geração paramétrica de capas de discos, presente neste projecto e no anterior, poderá constituir uma boa via de exploração futura no presente trabalho, tornando o sistema mais pessoal para o artista ou até para o consumidor da sua música.

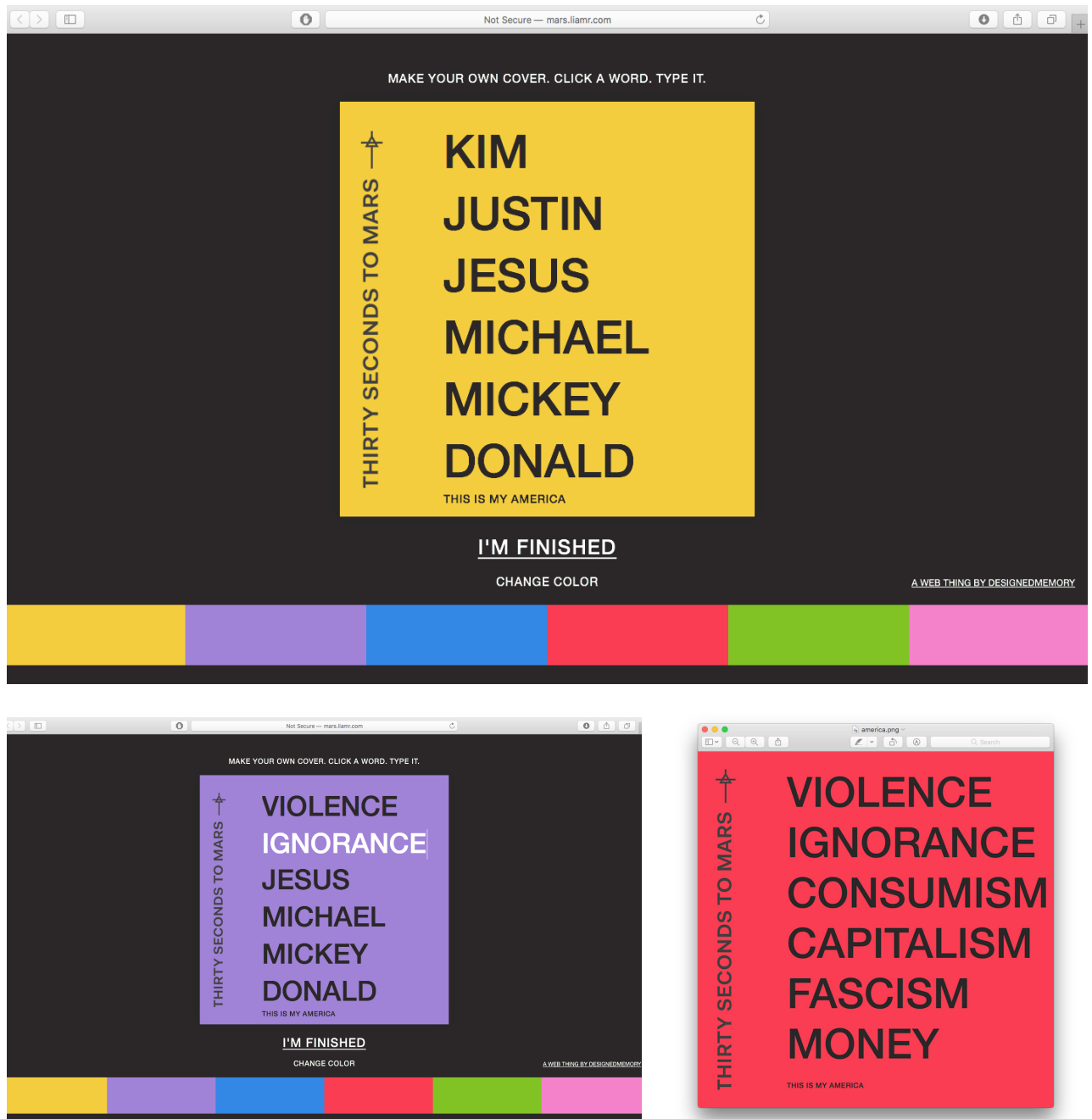


Figura 28 — Screenshots do processo de criação de uma capa personalizada do álbum *America* até ao output final em PNG.

53 America Cover Generator. <http://mars.liamr.com> (data de consulta: 17.06.2019)

3.6 Álbuns de música

Nesta secção apresenta-se três exemplos de álbuns de música que serviram de referência para a componente prática desta dissertação. Os critérios de selecção destes projectos foram **i)** a ideia da mensagem da música servir de motor para a criação gráfica, **ii)** a união que se construiu entre os diferentes materiais que se relacionam com o álbum de música e **iii)** a ligação criada entre a imagem gráfica do álbum de música e a identidade visual dos concertos. Todos estes aspectos foram tidos em conta no desenvolvimento da componente prática deste trabalho.

22, A Million, de Bon Iver, por Eric Timothy Carlson

A imagem gráfica deste álbum foi o motivo de inspiração para a ideação desta dissertação. Numa entrevista dada ao *Walker Art Center*⁵⁴, Eric Timothy Carlson — o designer que desenvolveu a imagem gráfica do álbum — referiu a importância dos símbolos ao longo de todo o processo. Carlson foi desenhando pequenos símbolos e rabiscos relativos a tudo o que ia acontecendo à sua volta, no estúdio, nas conversas que tinha com Justin Vernon, vocalista dos Bon Iver, etc.. No entanto, foi ao analisar as canções que os símbolos e anotações começaram a ganhar sentido de acordo com as letras (Fig. 29).

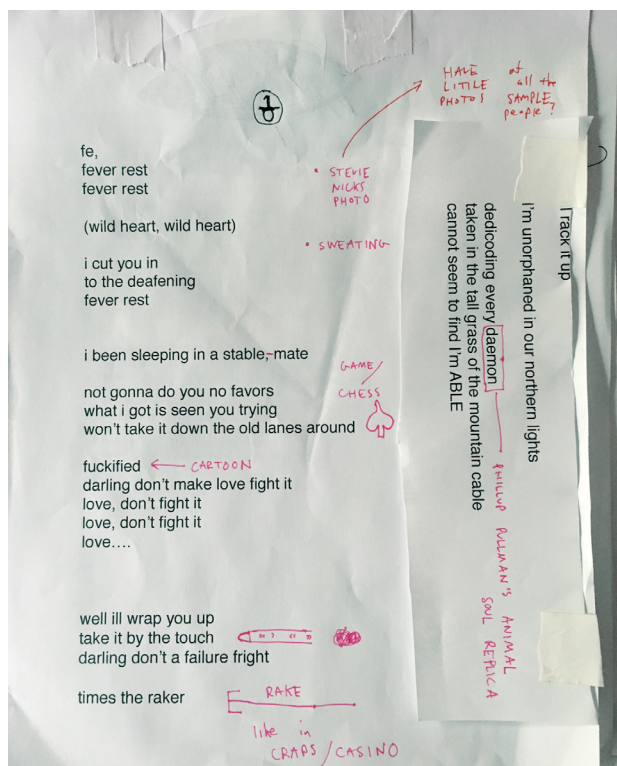


Figura 29 — Rabiscos feitos por Eric Timothy Carlson numa das letras.

54 Walker. <https://walkerart.org/magazine/designing-bon-iver-22-a-million-album-art-eric-timothy-carlson> (data de consulta: 17.01.2018)

No final do processo, cada canção tinha um conjunto associado de anotações e símbolos que, posteriormente, compuseram todo o grafismo do álbum⁵⁵. Este conceito originou a ideia da presente dissertação de desenvolver uma imagem gráfica para um álbum de acordo com as letras das músicas, tendo cada música uma identidade própria. Este último aspecto tornou-se mais claro quando Carlson desenvolveu os *lyric videos* de cada canção, todos diferentes uns dos outros, no entanto integrados no grafismo geral. A imagem gráfica foi aplicada e adaptada para todo o tipo de materiais (Fig. 30) — nomeadamente *outdoors*, *T-shirts*, latas de cerveja, etc. —, tornando-os reconhecíveis e identificativos do álbum, mesmo quando isolados. Este foi um dos aspectos que se teve em consideração durante as explorações para imagens gráficas realizadas neste projecto.

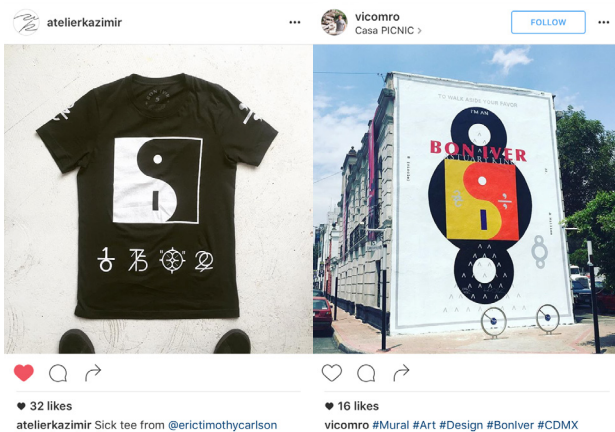
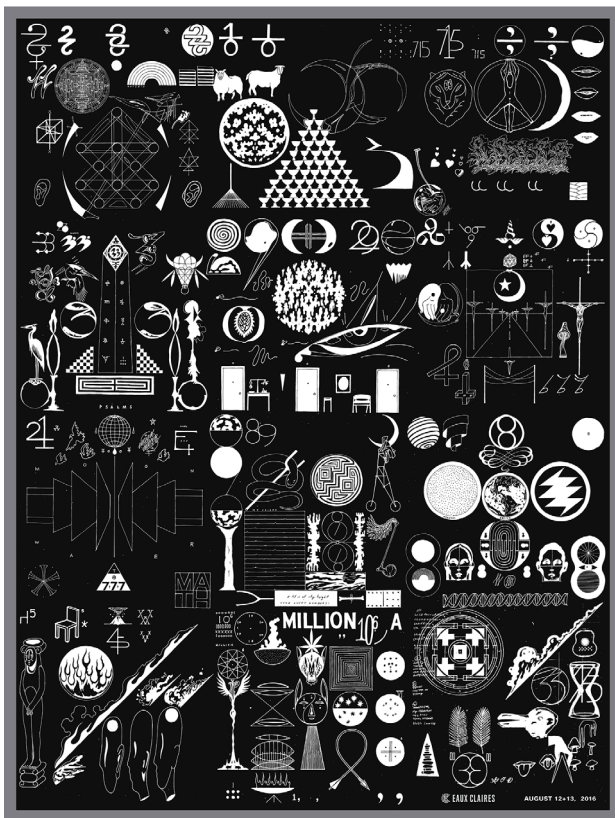


Figura 30 — Imagem gráfica do álbum 22, *A Million* dos Bon Iver, desenvolvida por Eric Timothy Carlson. Fica claro nestas imagens que o sistema gráfico funciona como um todo e cada peça funciona também isoladamente. É um bom exemplo de uma imagem gráfica para um álbum de música em que o que serve de base são as letras das músicas.

Awake, de Tycho

Tycho é o nome do projecto musical de Scott Hansen, músico e designer americano. Até este álbum, Hansen trabalhava nas duas áreas em separado, sob dois nomes diferentes — Tycho na música e ISO50 no design. No entanto, resolveu juntar ambas as áreas neste trabalho, tornando a imagem gráfica e os efeitos visuais dos concertos o reflexo gráfico da música que faz. Numa entrevista ao *The Verge*⁵⁶, Hansen fala da importância de, além da música, desenvolver também a componente visual do álbum. Este facto torna-se mais evidente quando se vê um dos seus espectáculos ao vivo (Fig. 31)⁵⁷: como deixa de ser necessário contratar alguém, que pode ou não entender a mensagem da música, para criar os efeitos visuais, torna-se tudo mais pessoal. Relativamente aos restantes materiais da imagem gráfica, Hansen diz ainda que, como vêm do mesmo sítio de onde a música vem, se tornam mais coesos e consistentes⁵⁸ (Fig. 32).



Figura 31 — Screenshot do vídeo do *live set* em San Francisco para o *Boiler Room*.



Figura 32 — À esquerda: *screenshots* do vídeo para a música *Montana*. À direita: algumas aplicações de objectos físicos.

⁵⁶ The Verge. <https://www.theverge.com/2014/2/25/5419816/tycho-interview> (data de consulta: 17.01.2018)

⁵⁷ YouTube. <https://youtu.be/vnkKnUwrTfy> (data de consulta: 17.01.2018)

⁵⁸ The Verge. <https://www.theverge.com/2014/2/25/5419816/tycho-interview> (data de consulta: 17.01.2018)

Este trabalho é mais um exemplo de como os visuais reflectem directa ou indirectamente a mensagem que a música pretende transmitir, mas de uma forma mais abstracta que o projecto descrito anteriormente. Neste aspecto, é uma boa referência para a presente dissertação, na qual se procura transmitir a mensagem e as emoções presentes na música, embora o método utilizado seja diferente, pois abrange uma componente automaticamente generativa.

A Head Full Of Dreams, dos Coldplay, por Pilar Zeta

Os Coldplay são o exemplo perfeito de uma banda que começou a desenvolver as suas composições musicais já a pensar em todo o espectáculo sonoro e visual dos concertos. Desde esta inversão na sua estratégia de marketing (os álbuns servem para promover os concertos e não o inverso), que as imagens gráficas desenvolvidas para os seus álbuns transparecem um ambiente de festa com cores garridas e composições elaboradas. No entanto, *A Head Full Of Dreams* é, provavelmente, o álbum da sua autoria em que existe mais ligação entre a imagem gráfica desenvolvida e os efeitos visuais dos concertos. A imagem gráfica é de Pilar Zeta e foi elaborada a partir de recortes e colagens (Fig. 33), com imagens dos elementos da banda quando eram crianças, a fim de remeter para sonhos — conceito central do álbum. No meio dos elementos, encontram-se também elementos presentes nas músicas e suas letras, como pássaros exóticos, planetas, etc.⁵⁹. No centro está representada a *Flower of Life*, e é este o símbolo, juntamente com as cores e a tipografia, que une todas as aplicações (Fig. 34) com os efeitos visuais dos concertos (Fig. 35)⁶⁰.



Figura 33 — Pilar Zeta durante a composição da colagem que daria origem à capa do álbum *A Head Full Of Dreams*.

59 Coldplay. <http://coldplay.com/interview-pilar-zeta-ahfod-artwork-creator/> (data de consulta: 17.01.2018)

60 YouTube. https://youtu.be/_kqbeUHYU6o (data de consulta: 17.01.2018)



Figura 34 — Algumas aplicações da imagem gráfica do álbum *A Head Full Of Dreams* dos Coldplay, desenvolvida por Pilar Zeta.



Figura 35 — Screenshots de um vídeo do concerto dos Coldplay em Buenos Aires, em 2017.

3.7 Conclusões

Neste capítulo foram abordados casos de estudo que foram importantes para a base da componente prática desta dissertação.

Na componente mais técnica, passou a conhecer-se abordagens possíveis para a obtenção de imagens a partir de texto e para o mapeamento entre características do som e outro tipo de variáveis. No que toca à geração de imagens a partir de texto, as abordagens mencionadas revelam-se bastante complexas, com recurso a algoritmos genéticos de inteligência

artificial. Procurou-se nesta dissertação desenvolver um método mais simples de obtenção de imagens, relativas a palavras-chave — substantivos e verbos — que sirvam de base às ilustrações. O estudo da música e das emoções e o mapeamento das suas variáveis com outras realizado por Mariana Seça na sua dissertação de Mestrado em Design e Multimédia (Seça, 2017) foi bastante relevante para o mapeamento das variáveis sonoras com as visuais presente nesta dissertação.

Nos casos de estudo que misturam a componente técnica com a gráfica abordou-se a ilustração generativa, tendo-se conhecido potenciais ferramentas para o desenvolvimento da componente generativa do sistema. O sistema inspira-se na implementação do projecto descrito, *Gold*, aliando-o à interacção com texto e som.

Também nesta parte se descreveu métodos de modelação de imagem de acordo com som com *outputs* gráficos bastante interessantes. O sistema desenvolvido neste trabalho corresponde a um mecanismo que mistura as capacidades descritas nos primeiros três projectos desta secção: **i)** obtenção de uma imagem estacionária modelada por som e que seja visualmente apelativa, a fim de poder ser utilizada no desenvolvimento de uma imagem gráfica para um álbum; **ii)** reacção da imagem a um som reproduzido em *real-time*, com o objectivo de gerar imagens também a partir de músicas pré-gravadas; **iii)** e reacção de efeitos visuais ao som gravado em *real-time*, de modo a que funcione em performances ao vivo. São descritos outros dois projectos que, embora não tenham servido de base para a implementação do trabalho que se apresenta, abrem portas para uma possível exploração futura desta dissertação no contexto da tipografia, complementando ou substituindo as ilustrações pela própria letra da música, ou, ainda, no contexto da *web*, permitindo um meio de interacção artista-público.

Por fim, encerra-se o capítulo com a apresentação de exemplos de imagens gráficas para álbuns de música que retratam bem o que se pretende com o resultado da segunda fase deste projecto: criação da imagem gráfica para um álbum de música que una o objecto artístico com os materiais e eventos associados. O conceito do primeiro projecto descrito assemelha-se ao da presente dissertação, no entanto, todos os símbolos são criados manualmente e sem qualquer automatismo, ao contrário do que acontece neste trabalho. Além disso, a aplicação da imagem gráfica não se estende aos efeitos visuais dos concertos, que é uma das problemáticas que se pretende solucionar com este trabalho. O segundo e o terceiro projectos desta última secção, resolvem bem esta questão, no entanto, sempre com técnicas manuais e não automáticas.

Metodologia

Capítulo IV

- 4.1 Objectivos
- 4.2 Processo
- 4.3 Plano de trabalho

Capítulo IV

Metodología

A indústria da música tem sido particularmente sensível à evolução tecnológica, quer ao nível da produção musical, quer ao nível da sua distribuição ou até mesmo dos concertos. A tecnologia tem contribuído para oferecer aos artistas e ao público novas plataformas, novos recursos, novas experiências para criação ou fruição da música. Este projecto alinha-se com esta ideia de que a tecnologia pode contribuir para melhor produção ou melhor fruição da música, potenciando a interpretação quer dos sons, quer da mensagem contida na letra da música, ao mesmo tempo que permite a criação de conceito de álbum de música associado a essa interpretação.

A criação de um sistema que possa gerar, em modo semi-automático, imagens associadas à letra da música e ambientes gráficos que respondem de acordo com características do som a ela associado é uma área relativamente nova que se pretendeu explorar. Existem vários casos de estudo que serviram de base de inspiração, mas pensa-se que há um mundo de potencialidades a descobrir nesta relação do mundo da música e da tecnologia. Optou-se então por uma abordagem exploratória, procurando conhecer de modo aprofundado o domínio de ação, procedendo com base na tentativa-erro, com algumas reformulações, até se atingir um produto final que se considera ajustado à resposta a dar ao problema em causa.

Poderá então afirmar-se que o desenvolvimento deste projecto se orientou pela metodologia de *Design Science Research*, uma vez que esta se baseia na procura de soluções para um dado problema, partindo de um conhecimento aprofundado do domínio em que se insere o problema para depois, em processo iterativo, ir estudando e encontrando soluções, analisando o seu desempenho até encontrar um produto final que responda de modo ajustado ao problema (Lacerda, 2013).

4.1 Objectivos

O objectivo central desta dissertação é a criação de um sistema que gere ilustrações a partir de letras de músicas e que as molde de acordo com o som, de modo a obter um *output* que permita transparecer as mensagens e emoções contidas numa música. Pretende-se que esse *output* seja capaz de ser utilizado como base de exploração para o desenvolvimento de uma imagem gráfica para um álbum de música e dos efeitos visuais dos concertos correspondentes, de forma a enaltecer a função do *output* do sistema.

Para atingir este objectivo, foi necessário estabelecer e cumprir diversos objectivos específicos:

- i) Pesquisar sobre a história da indústria fonográfica e da música ao vivo até à actualidade, assinalar as problemáticas presentes em cada momento, procurar as soluções já propostas para cada uma delas e, por fim, situar a dissertação numa problemática actual.
- ii) Encontrar uma estrutura funcional para o sistema, de modo a que seja possível o seu *output* transparecer visualmente as mensagens e emoções contidas em qualquer música.
- iii) Encontrar uma forma de analisar a mensagem contida na letra da música.
- iv) Implementar um programa que utilize uma imagem como molde para gerar ilustrações visualmente apelativas, mantendo a sua legibilidade.
- v) Implementar um programa que manipule as ilustrações geradas e adapte o ambiente gráfico de acordo com características do som, mantendo um aspecto visual apelativo.
- vi) Desenvolver o sistema de modo a obter *outputs* estáticos e animados, de modo a que resultem graficamente em ambos os formatos.
- vii) Explorar formas de desenvolvimento de uma imagem gráfica para um álbum de música, para que esta seja apelativa e que saliente os objectivos do sistema: ilustrar visualmente as mensagens e emoções contidas nas músicas que constituem o álbum.
- viii) Utilizar o sistema como gerador de efeitos visuais para um concerto, com a finalidade de criar um ambiente imerso e visualmente apelativo que transpareça as mensagens e emoções contidas na performance das músicas

4.2 Processo

O processo deste trabalho decorreu em duas componentes, uma teórica e outra prática.

Componente teórica

Esta dissertação sustenta-se numa base teórica construída durante a pesquisa sobre as diversas temáticas presentes na componente prática. A partir da leitura de livros, artigos e várias páginas da Internet, obteve-se o fundamento teórico do estado da arte, que contextualiza toda a dissertação a nível histórico e da actualidade.

Numa fase mais específica da investigação para o estado da arte, são exploradas diversas temáticas abrangidas na dissertação, tendo por base casos de estudo actuais. A pesquisa por projectos relacionados com

a dissertação permite conhecer um pouco do que tem sido feito na área, analisando pontos fortes e fragilidades de cada um.

Componente prática

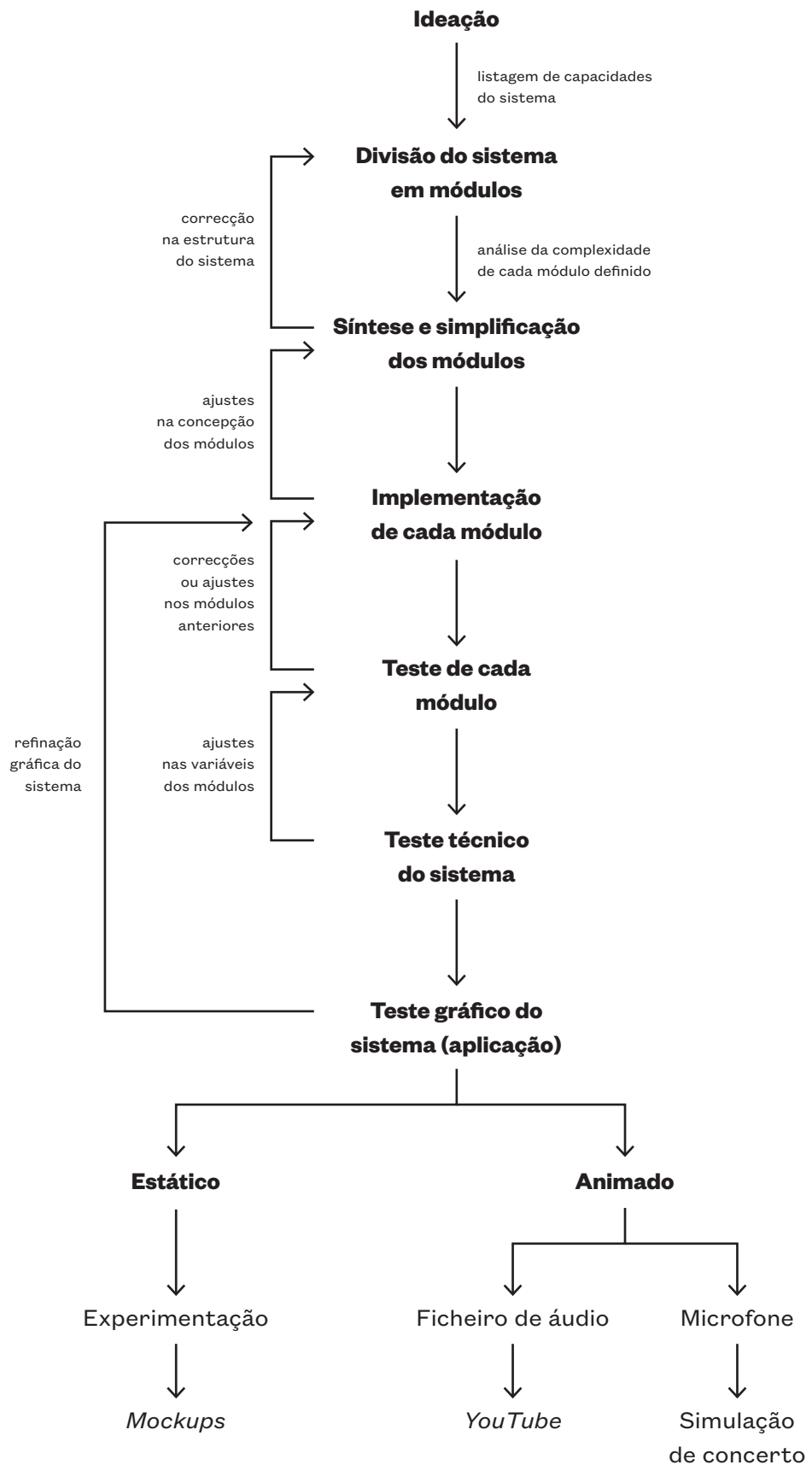
Na componente prática são aplicados, nos diversos processos necessários para a realização do projecto, os conhecimentos adquiridos durante a pesquisa. Esta componente desenvolveu-se em duas fases: a primeira (**fase 1**) consiste no desenvolvimento de um sistema de geração de imagens a partir de letras de músicas, a modelação das imagens e adaptação do ambiente gráfico de acordo com características do som; a segunda (**fase 2**) trata de uma demonstração de possíveis aplicações do *output* deste sistema **i**) numa imagem gráfica para os materiais relacionados com um álbum de música e **ii**) nos efeitos visuais dos concertos correspondentes.

A **fase 1** partiu de um *brainstorming* que teve como objectivo desenvolver uma ideia para o funcionamento do sistema. Para facilitar este processo, realizou-se uma enumeração das capacidades que se pretendia inicialmente que o sistema tivesse. Tendo esta lista de capacidades, foi mais fácil dividir o sistema em módulos dedicados a cada uma, de modo a que pudessem funcionar em cadeia até atingir um *output* final. De seguida, passou-se para uma fase de síntese e simplificação de alguns dos módulos, decidindo-se tornar o sistema semi-automático em vez de completamente automático, atendendo a que a performance do *output* final não fosse prejudicada no seu funcionamento em *real-time*. Começou-se por seleccionar uma canção para testar e implementar cada um dos módulos pela sua ordem na cadeia do sistema. Quando se terminava a implementação de cada módulo, partia-se para o seguinte, voltando ao anterior se se detectasse algum erro ou fragilidade no seu funcionamento que prejudicasse o seguinte. No final, todo o sistema foi testado, a fim de confirmar se tudo funcionava como se previu inicialmente, ou seja, se, ao correr o **módulo #5** do sistema, **i**) as ilustrações estavam sincronizadas com as palavras respectivas da letra, **ii**) as variações visuais correspondiam às variações do som e **iii**) o ambiente gráfico do *output* final era visualmente diferente em canções com emoções distintas. Ajustaram-se algumas variáveis do programa até que o *output* satisfizesse estas condições e corrigiram-se todos os erros detectados durante estes testes.

Seguiu-se depois para a **fase 2**, ou seja, para o processo de aplicação do sistema, com o fim de testar o funcionamento do seu *output* a nível gráfico. Nesta fase, desenvolveram-se diversas capas de disco para uma canção seleccionada, aplicando-as em *mockups*, para que o resultado visual fosse mais realista. Testou-se também o seu funcionamento como *output* animado, tanto na versão do sistema que se baseia na leitura de um ficheiro de áudio, como na versão que se baseia no *input* do microfone. Para o primeiro, experimentou-se colocar o vídeo gerado no *YouTube*, uma das principais plataformas para visualizar e difundir videoclipes, com o objectivo de testar o *output* no contexto em que se pretende que seja utilizado. Para o segundo, simulou-se um cenário de concerto, utilizando um projector ligado ao

computador com o sistema a correr em *fullscreen*, de modo a servir de fundo à performance da mesma canção.

Todos estes processos organizam-se de acordo com o seguinte mapa:



4.3 Plano de trabalho

O plano de trabalho definido inicialmente teve em conta um ano lectivo, no entanto, a meio do ano, a proposta teve de ser refinada, de modo a excluir tarefas que se revelaram demasiado ambiciosas e a rever a estrutura do sistema, a fim de cumprir o mesmo propósito, mas de modo mais acessível. Como tal, novas tarefas surgiram no plano, sendo certo que, para algumas, foi necessário reservar mais tempo do que o que estava previsto inicialmente, nomeadamente para o estudo aprofundado das características do som e da sua eventual ligação com emoções. Para tal, foi preciso explorar o *Pure Data*. Esta exploração incluiu a frequência opcional na unidade curricular de Design de Som, e a realização de alguns projectos fora do âmbito curricular, como a imagem gráfica do festival *Criatek 2018* — um trabalho onde os conhecimentos adquiridos sobre o som e o *Pure Data* foram aplicados e aprofundados, tanto na análise, como na produção de som, de modo a criar um ambiente gráfico associado a características sonoras. Desta forma, tendo identificado, de modo mais objectivo, as necessidades de exploração, reformulou-se o plano de trabalho, integrando as novas tarefas e estendendo o tempo de exploração para mais um ano de trabalho. Assumiu-se que o período de desenvolvimento deste projecto era um período de aprendizagem particularmente importante, em que se desejava aprofundar e explorar o conhecimento para desenvolver o sistema sem restrições significativas de tempo.

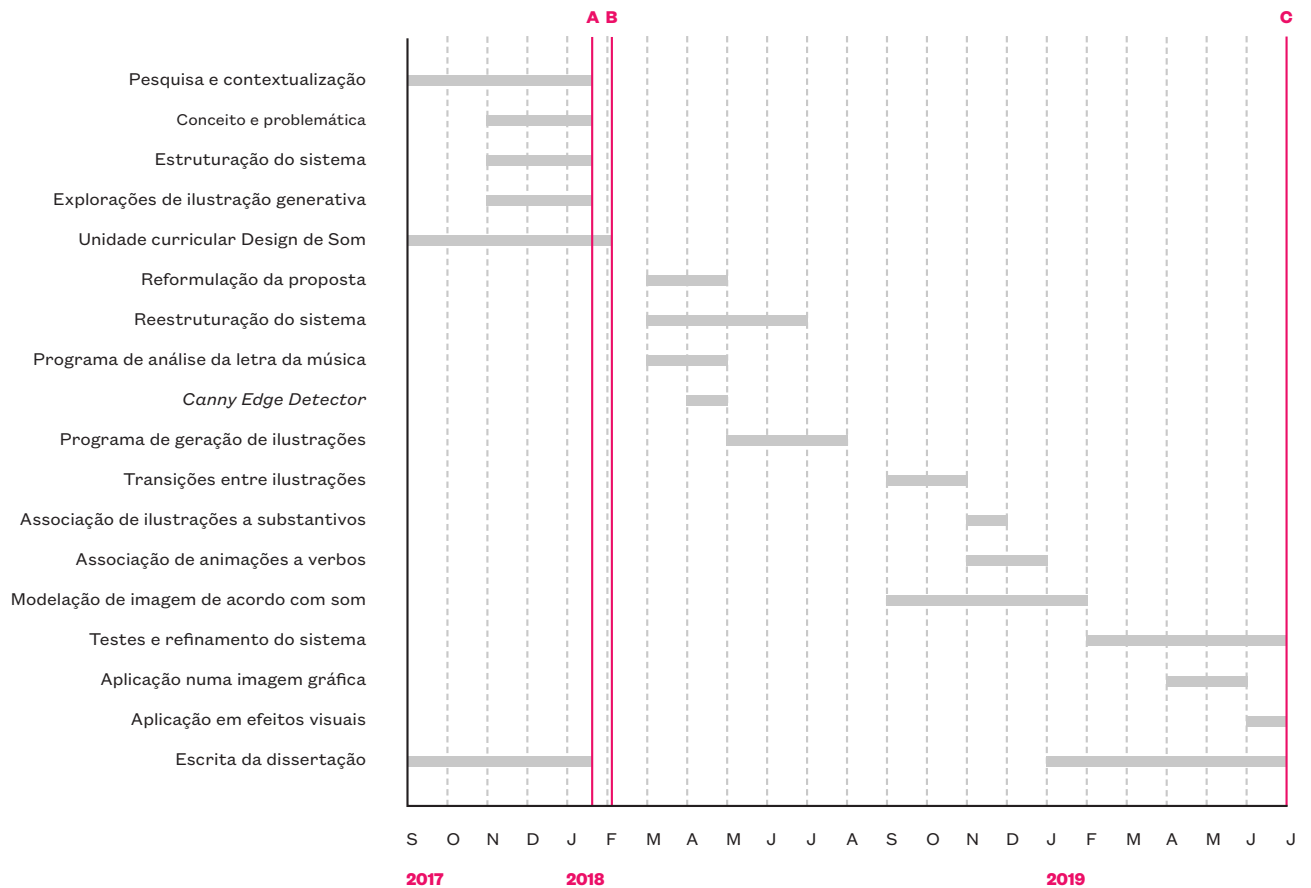
O plano de trabalho inclui diversas tarefas desenvolvidas ao longo do tempo, que se enumeram de seguida:

1. Pesquisa e contextualização histórica
Setembro 2017 — Janeiro 2018
2. Definição do conceito e problemática
Novembro 2017 — Janeiro 2018
3. Estruturação do sistema a desenvolver
Novembro 2017 — Janeiro 2018
4. Exploração de métodos para ilustração generativa a partir uma imagem de base
Novembro 2017 — Janeiro 2018
5. Frequência opcional da unidade de curricular Design de Som
Setembro 2017 — Fevereiro 2018
6. Apresentação da proposta de trabalho
Fevereiro 2018
7. Entrega do projecto de Design de Som
Fevereiro 2018
8. Reformulação da proposta de trabalho
Março 2018 — Abril 2018
9. Reestruturação do sistema a desenvolver
Março 2018 — Junho 2018
10. Desenvolvimento do programa de análise da letra da música
Março 2018 — Abril 2018

84 Metodologia

- 11.** Aplicação do *Canny Edge Detector* do *OpenCV* para filtrar a imagem base
Abril 2018
- 12.** Desenvolvimento do programa que gera ilustrações a partir de uma imagem base
Maio 2018 — Julho 2018
- 13.** Exploração de transições entre as ilustrações geradas
Setembro 2018 — Outubro 2018
- 14.** Desenvolvimento da componente do programa que associa uma ilustração a um substantivo
Novembro 2018
- 15.** Desenvolvimento da componente do programa que associa uma animação a um verbo
Dezembro 2018
- 16.** Desenvolvimento da componente do programa que molda o aspecto gráfico de acordo com características do som
Setembro 2018 — Janeiro 2019
- 17.** Testes e refinamento do sistema
Fevereiro 2019 — Junho 2019
- 18.** Aplicação do sistema na exploração para o desenvolvimento de uma imagem gráfica para um álbum de música
Abril 2019 — Maio 2019
- 19.** Aplicação do sistema nos efeitos visuais para um concerto
Junho 2019
- 20.** Escrita da dissertação
Setembro 2017 — Janeiro 2018 e Janeiro 2019 — Junho 2019
- 21.** Entrega final
Julho 2019

Diagrama de Gantt



- Metas**
- A** Entrega intermédia
 - B** Entrega do projecto de Design de Som
 - C** Entrega final

Projecto

Capítulo V

- 5.1 Descrição do projecto
- 5.2 Módulo #1
- 5.3 Módulo #2
- 5.4 Experiências preliminares
de ilustração generativa
- 5.5 Módulo #3.1
- 5.6 Módulo #3.2
- 5.7 Módulo #4
- 5.8 Módulo #5
- 5.9 Atributos sonoros estudados
mas não utilizados
- 5.10 Módulo #6
- 5.11 Módulo #7
- 5.12 Discussão de resultados

Capítulo V

Projecto



Este capítulo inicia-se com a definição do conceito do projecto, passando à descrição do funcionamento do sistema desenvolvido. Posteriormente, são descritos em detalhe os diferentes módulos que integram o sistema.

O foco desta dissertação — explorar o conceito de álbum como mensagem e objecto artístico — pressupõe o reconhecimento de que as maiores mudanças na indústria da música aconteceram desde o aparecimento da Internet e de que há ainda muito a explorar nesta área até encontrar estratégias de design e de marketing que funcionem, em simultâneo, a nível artístico e económico na era digital em que vivemos. Pressupõe também a consciência de que existe um domínio dos *singles* em relação aos álbuns conceptuais.

Também no foco desta dissertação está a forma como a indústria da música ao vivo tem contribuído para a integridade do conceito de álbum. De facto, o paradigma anterior à era digital, em que os concertos serviam de promoção para os álbuns de música lançados, inverteu-se e, actualmente, os álbuns são já concebidos com os concertos em mente, tanto a nível de composição musical, como a nível de design e estratégia de marketing. Toda esta contextualização histórica situa a presente dissertação na indústria musical actual, salientando os problemas para os quais se explorou estratégias e abordagens alternativas que permitissem minimizá-los: **i)** sente-se nos dias de hoje que a mensagem contida nas letras das músicas que compõem um álbum de música é menos valorizada, tanto pelo artista no seu processo de criação, como pelo consumidor quando a ouve, pois o que se procura e divulga hoje em dia são músicas fáceis de ouvir, de consumo rápido e com uma mensagem clara e fácil de interpretar; **ii)** nesta era digital, o mercado musical está muito mais presente no meio digital do que no meio físico, mantendo, no entanto, alguma força nos objectos de colecção, pelo que as imagens gráficas dos álbuns de música exigem alguma versatilidade, qualidade e capacidade de ligação entre os dois meios; **iii)** os concertos também exigem que exista uma ligação visual com o álbum de música que representam, sendo certo que, com o crescimento dos festivais de música, em que se misturam diversas bandas com identidades gráficas diferentes, os concertos individuais necessitam de oferecer espectáculos audiovisuais cada vez mais inovadores e únicos, de modo a conquistar os consumidores dos festivais — afinal, num festival existe uma relação quantidade-preço mais favorável do que num concerto individual, porém, a relação qualidade-preço poderá ser a grande diferença, se o espectáculo for bem pensado e construído.

5.1 Descrição do projecto

Partindo dos problemas que deram origem à presente dissertação, o projecto construiu-se com a intenção de desenvolver um método para gerar imagens a partir das mensagens e emoções presentes num álbum de música, tendo como finalidade possibilitar a criação de uma imagem gráfica para todos os materiais e eventos associados a esse conjunto de músicas, oferecendo-lhe integridade conceptual e visual.

Para tal, desenvolveu-se um sistema semi-automático que produz imagens a partir da letra de uma música à escolha do utilizador e as transforma de acordo com alguns atributos sonoros dessa música. Estas imagens podem ser utilizadas para **i**) criar os materiais gráficos do *single* ou do álbum (se o sistema for aplicado a todas as faixas do álbum), **ii**) produzir vídeos para cada canção do álbum (podendo a mesma canção ter vários vídeos diferentes, dependendo das ilustrações seleccionadas pelo programa) e **iii**) efeitos visuais para os concertos relativos ao álbum. Desta forma, todos os materiais e eventos relacionados com esse álbum de música estarão integrados num só sistema gráfico.

No desenvolvimento do projecto, foram inicialmente identificados dois elementos presentes numa música convencional: letra e som. A integração do sistema foi feita de modo a que **i**) a letra servisse de motivo de inspiração para a geração de ilustrações, e **ii**) as características sonoras da música moldassem essas ilustrações, obtendo-se como produto final um ambiente gráfico que poderá induzir a emoções. A Tabela 1 apresenta uma relação entre a função de cada elemento de uma música identificado e a respectiva representação visual criada pelo sistema.

Elementos	Função	Analogia visual
Letra	Contém a mensagem que se quer transmitir para o público.	Ilustrações geradas pelo sistema.
Som	Define uma melodia e um ambiente global para a letra da música que induzem a emoções.	Modelação das ilustrações e adaptação do ambiente gráfico geral.

Tabela 1 — Analogia entre a função dos elementos da música e as acções do sistema.

Componente #1 — Letra

A letra da música é o foco principal do sistema, pois é a portadora da mensagem que se pretende transmitir para o público. Assim, procurou-se, a partir dela, encontrar palavras-chave (substantivos e verbos) que servissem de ponto de partida para a exploração de imagens ou para a criação de ilustrações, no caso dos substantivos, e para a animação das ilustrações, no caso dos verbos. Explorou-se uma forma visualmente atraente de criar ilustrações e transições fluidas entre si, de modo a que o sistema funcionasse

não só em modo estático, como também em movimento. Desta forma, foi possível, numa segunda fase do projecto, aplicar o sistema na utilização de ilustrações para os materiais estáticos, no desenvolvimento de materiais de vídeo e na apresentação dos efeitos visuais dos concertos.

Componente #2 — Som

O som numa música apresenta determinadas características que permitem atribuir emoções à mensagem contida na letra da música. Esta componente do projecto pretende estabelecer uma analogia a esta função, modelando as ilustrações de acordo com as características do som. Para isso, foi necessário conhecer e seleccionar características do som, para posteriormente mapeá-las a alguns atributos da ilustração e do ambiente gráfico.

O sistema foi desenvolvido com o objectivo de permitir **i)** fazer uma leitura visual subjectiva da mensagem contida na letra da música, **ii)** interpretar algumas emoções presentes na música e **iii)** representar visualmente o ambiente sonoro e emocional no qual a mensagem se insere.

É possível aplicar o *output* do sistema em materiais estáticos — imagem gráfica para os materiais físicos relativos a um *single* ou álbum conceptual de música — ou animados — videoclipes ou efeitos visuais para concertos.

Aplicação #1 — Estático

O sistema desenvolvido permite a criação de uma imagem gráfica com o objectivo de **i)** transparecer mensagens e emoções presentes num *single* ou num álbum de música, **ii)** unificar visualmente todos os elementos pertencentes ou relacionados com o *single*/álbum, sendo reconhecíveis como pertencentes ao álbum quando isolados e **iii)** despertar a atenção do público para que procure conhecer a música.

Aplicação #2 — Animado

O sistema desenvolvido está desenhado de modo a **i)** ler ficheiros de áudio, podendo, assim, gerar um videoclipe, ou **ii)** receber som de um microfone, gerando em *real-time* efeitos visuais para um concerto.

O sistema integra sete módulos que funcionam em cadeia: **#1** análise e preparação da letra da música, **#2** definição dos tempos das palavras-chave ao longo da música, **#3** recolha de imagens, aplicação do filtro *Canny Edge Detector* e aprimoramento das imagens recolhidas quando necessário, **#4** ilustração e animação com base nas imagens filtradas, **#5** mapeamento de atributos sonoros com variáveis visuais, **#6** aplicação do sistema na imagem gráfica de um álbum de música e **#7** nos efeitos visuais dos respectivos concertos. O sistema é descrito como semi-automático, pois nos primeiros três módulos é necessária a participação do utilizador para preparar os materiais que serão posteriormente utilizados automaticamente pelos dois módulos seguintes. Estes dois módulos concentram-se num só programa,

no entanto, serão descritos em separado. Já o módulo #6 é onde o utilizador pode aplicar o *output* gerado pelo sistema no seu *single* ou álbum, podendo utilizar o seu próprio registo dentro de uma abordagem à sua escolha. Por fim, o último módulo é uma versão do módulo #5 para funcionamento *real-time*, no sentido em que o *input* sonoro do sistema é um microfone em vez de um ficheiro de uma música. Para efeitos de descrição e demonstração, considere-se que se está a utilizar este sistema para criar a imagem gráfica de uma faixa musical isolada, ou seja, um *single* — para ser aplicado num álbum de música, basta repetir todo o processo que será descrito de seguida em todas as faixas do álbum. Desta forma, seleccionou-se uma canção da minha autoria: *Take Me From Here*.

5.2 Módulo #1 — Análise e preparação da letra da música

Este módulo é responsável por analisar a letra do *single* e recolher palavras-chave que possam ser posteriormente utilizadas pelo utilizador e pelo programa. Uma letra de uma música consiste basicamente num texto, ou seja um conjunto de palavras que, quando submetidas a determinados mecanismos de coerência e coesão textual, formam um todo com sentido. No entanto, existem muitas palavras presentes num texto que não são significantes, ou seja, não têm uma representação visual, como, por exemplo, palavras que apenas são utilizadas como elo de ligação entre ideias — é o caso dos conectores ou das preposições —, ou para ajudar a construir o valor referencial dos objectos — por exemplo, os artigos. Assim, optou-se por filtrar do texto apenas as palavras com possíveis representações visuais: substantivos, verbos, adjectivos, advérbios e interjeições. A fim de simplificar o sistema, decidiu-se utilizar apenas substantivos e verbos para representar visualmente. Tendo em conta que os substantivos representam objectos (os substantivos são significantes e os objectos são o significado), considerou-se que os substantivos deveriam resultar em ilustrações relativas ao seu significado, e, como os verbos representam acções — normalmente são eventos que ocorrem durante um período de tempo, pelo que, muitas vezes, é difícil representar visualmente numa imagem estática —, concluiu-se que deveriam corresponder a animações, no caso, na animação das ilustrações relativas aos substantivos. Para tal, utilizou-se a biblioteca *RiTa* do *Processing*, que oferece diversas funções para analisar e tratar texto. Tendo por base um ficheiro *TXT* com a letra do *single* seleccionado, preparado pelo utilizador, utiliza-se a função `pos()` para obter a *part of speech* (POS) de cada palavra da letra. Por exemplo, na frase “He ate the apple”, a função `pos()` retorna `prp vbd dt nn`, o que, segundo a *PENN tag set*, significa: “personal pronoun”, “verb, past tense”, “determiner” e “noun, singular or mass”⁶¹. Desta forma, é possível obter a POS de cada palavra e seleccionar apenas as palavras que são relevantes, neste caso, as que

61 RiTa Reference. <https://rednoise.org/rita/reference/RiString/pos/index.php> (data de consulta: 30.06.2019)

pertencem às classes de palavras filtradas, para, posteriormente, guardar numa tabela juntamente com o seu POS. Para este trabalho guardou-se apenas os substantivos, os verbos e os adjectivos (que, apesar de não serem ainda utilizados pelo sistema, a sua utilização poderá vir a ser considerada em trabalho futuro). Algumas palavras estavam associadas a POS que, no contexto, eram incorrectas, pelo que foi necessário corrigir manualmente. Por exemplo, a palavra “move”, presente no verso “But I can not move on” do single seleccionado, foi identificada como substantivo, embora, no contexto do verso, se trate de um verbo. Esta correcção foi importante, pois a palavra “move” faz parte das quatorze acções básicas que serão explicadas mais à frente.

Para a presente abordagem, os substantivos só têm de ser guardados na tabela para que o utilizador (ou o sistema, se essa funcionalidade vier a ser desenvolvida no futuro) possa consultá-los com o objectivo de pesquisar ou criar imagens que os ilustrem. No entanto, o processamento dos verbos é mais complexo, pois servem de base para as animações das imagens. Assim, foi necessário reflectir sobre como seria possível criar de forma automática uma animação concordante com qualquer verbo que surgisse na letra. Apresenta-se abaixo a abordagem escolhida para o processamento e utilização dos verbos neste sistema.

No seu projecto “The Shape of Story”, Duri Long (Long *et al.*, 2017) definiu quatorze acções básicas adaptadas a partir do estudo de Roger C. Schank (Schank, 1973), que defende que todos os verbos existentes podem ser mapeados para quatorze acções básicas. As acções básicas definidas por Duri Long são: *smell, see, hear, feel, conclude, think, be, speak, move, transport, ingest, have, propel* e *expel*. Por exemplo, os verbos *go, drive, run* e *walk* poderiam ser todos mapeados para a acção básica *transport*. No trabalho de Duri Long, a representação visual de cada verbo corresponde à da acção básica para a qual o verbo é mapeado e trata-se de símbolos desenhados matematicamente e inspirados nas pinturas de Joan Miró e nos hieróglifos egípcios.

Tendo em conta que o presente projecto requer um funcionamento não só estático, mas também dinâmico, considerou-se que era importante utilizar animações, pelo que se aproveitou a utilização dos verbos das letras — que correspondem maioritariamente a acções — para criar essas animações, utilizando a mesma estratégia que Duri Long utilizou para mapear os verbos do texto aos símbolos: cada verbo que surge na letra da música é mapeado para uma acção básica à qual corresponde uma animação. Para tal, agruparam-se as quatorze acções básicas em seis grupos, associando a cada um uma animação ou efeito visual. Este mapeamento entre acção e representação gráfica foi a solução escolhida e implementada para esta componente. Embora outras possíveis abordagens pudessem ter sido exploradas, optou-se por esta, uma vez que se revelou uma solução que se enquadrava no âmbito deste projecto. Abaixo apresenta-se o agrupamento efectuado de acordo com a interpretação pessoal de cada acção básica, como esta se relaciona com as restantes e ainda a correspondente animação pensada inicialmente. No módulo #4, faz-se uma comparação entre estas animações e as finais.

Grupo #1 — *smell, see, ear, feel*

Associado aos cinco sentidos.

Animação inicial: mover os pontos da ilustração ao longo de uma circunferência, pois os cinco sentidos são muitas vezes esquematizados num círculo. Além disso, a cabeça, no qual estão presentes a maioria dos sentidos, é redonda, pelo que o círculo parece ser um símbolo representativo deste grupo.

Grupo #2 — *conclude, think, be*

Associado à filosofia/pensamento.

Animação inicial: colocar um *glow* nos pontos da ilustração, ou alternar entre o desfoque e a nitidez dos pontos para dar a ideia de um pensamento a ganhar forma.

Grupo #3 — *speak*

Associado ao som da voz.

Animação inicial: atribuir uma vibração aos pontos da ilustração, pois o som é gerado a partir de vibração.

Grupo #4 — *move, transport*

Associado a movimento.

Animação inicial: espalhar os pontos aleatoriamente e voltar a colocá-los no lugar, para que eles se movam pelo ecrã.

Grupo #5 — *ingest, have*

De fora para dentro.

Animação inicial: concentrar os pontos no centro e voltar a colocá-los no lugar.

Grupo #6 — *propel, expel*

De dentro para fora.

Animação inicial: espalhar os pontos para fora e voltar a colocá-los no lugar.

A fim de detectar qual a acção básica que melhor representa o significado do verbo actual, está a ser utilizada, de momento, a função `getDistance()` do *WordNet*, a partir da biblioteca *RiTa* do *Processing*. Esta função permite calcular a distância mínima entre duas palavras da mesma POS na árvore da *WordNet*, obtendo um valor normalizado entre 0 e 1. A *WordNet* é composta por palavras que são dispostas numa árvore de acordo com a relação de sinonímia entre si. Assim, na função `getDistance()`, a distância devolvida corresponde à distância mínima entre as palavras e o primeiro nó que for

parente comum das duas, calculada em número de nós percorridos. Se este nó não existir, ou seja, se as palavras não tiverem qualquer relação de sinonímia, é devolvido o valor 1.0, que representa a distância máxima. Sempre que surge um verbo na letra da música, o programa calcula a distância entre esse verbo e cada uma das quatorze acções básicas. A acção básica com o valor mais baixo devolvido pela função `getDistance()`, ou seja, a que tem o significado mais próximo do do verbo em questão, define o grupo em que o verbo se insere e, portanto, a animação a ele associada.

No projecto de Duri Long, é utilizado o *Word2Vec* (W2V) — um modelo de *word embedding*, utilizado para atribuir números a palavras a partir de *machine learning*, a fim de poder trabalhar com elas em forma de vectores e que, conseqüentemente, possibilita o cálculo da distância entre palavras de acordo com determinado atributo, neste caso a sinonímia —, para atingir um resultado semelhante ao que se pretende neste projecto (Fig. 36). Este modelo aparenta ter resultados mais satisfatórios no cálculo da distância entre um verbo e cada uma das quatorze acções básicas, pois tem um funcionamento bastante diferente do da *WordNet* (Fig. 37) em diversos aspectos que são importantes para o resultado obtido em cada um deles. Ao comparar os dois modelos, concluiu-se que:

- i) os valores obtidos na função `getDistance()` são muito menos precisos do que os valores obtidos com o modelo W2V: ao experimentar-se a secção *Similarity of two words* de uma demonstração do W2V criada pelo Turku NLP Group⁶², percebeu-se que os valores devolvidos pelo modelo correspondem ao grau de similaridade entre duas palavras e vão de -1 (palavras opostas) até 1 (a mesma palavra), com uma precisão de sete casas decimais, enquanto os da função da biblioteca *RiTa* correspondem à distância em número de nós percorridos na *WordNet* entre duas palavras e, após normalização, vão de 0 (palavras muito próximas) até 1 (palavras muito distantes), com uma precisão de, no máximo, duas casas decimais.
- ii) com o modelo W2V é possível dispor todas as palavras num espaço de duas dimensões de acordo com o seu significado, enquanto o segundo dispõe as palavras numa árvore de acordo com a relação de sinonímia entre elas. A distância entre as palavras obtida com o W2V, portanto, **i)** simboliza a distância entre os seus significados, **ii)** as palavras nunca se sobrepõem totalmente (o que significa que o valor 1.0 só é retornado quando o verbo e a acção básica correspondem exactamente à mesma palavra) e **iii)** retorna sempre uma distância, independentemente de as palavras terem alguma relação entre elas ou não. Na função `getDistance()`, a distância devolvida corresponde à distância mínima entre as palavras e o primeiro nó que for parente comum das duas, calculada a partir do número de nós percorridos. Se

a palavra *moving* passa para *move* e a palavra *birds* passa para *bird*. Ainda assim, esta estratégia não funcionou para todos os casos. Um exemplo é o verbo *made* que passou pelas duas funções sem alterar para *make*, pelo que esta mudança teve de ser feita manualmente.

A fim de ilustrar o que foi descrito até agora, considere-se a frase:

The rock is rolling too fast.

O programa desenvolvido neste módulo funciona da seguinte forma:

- i) A palavra *the* é ignorada, pois é classificada como um artigo.
- ii) A palavra *rock* é guardada na tabela como substantivo, a fim de, posteriormente, ser criada uma ilustração, como será descrito nos módulos #3 e #4.
- iii) O verbo *is* é mapeado para a acção básica *be* e, portanto, atribui a animação associada ao grupo #2 (efeito de *glow* nos pontos) à ilustração do substantivo anterior, neste caso *rock*.
- iv) O verbo *rolling* é conjugado no infinitivo e é, de seguida, mapeado para a acção básica *move*, atribuindo, assim, a animação correspondente ao grupo #4 (espalhar dos pontos) à ilustração do substantivo anterior, que é, ainda, a palavra *rock*. Neste caso, tendo em conta que o verbo *rolling* surge logo a seguir ao verbo *is*, provavelmente, não existe tempo suficiente para a animação do verbo *is* ocorrer até ao fim, sendo, então, interrompida para dar lugar à animação do verbo *rolling*. Ao experimentar combinar duas animações diferentes concluiu-se que podem surgir problemas: quase todas as animações funcionam pelo cálculo das posições dos pontos e esse cálculo pode ser oposto em duas animações diferentes — como acontece nas animações dos grupos #5, em que os pontos se deslocam para o centro, e #6, em que os pontos se deslocam para fora — ficando os pontos na mesma posição.
- v) As palavras *too* e *fast* são, também, ignoradas, pois são classificadas como advérbios.

A Figura 38 apresenta a tabela resultante do programa.

POS	Original Word	Word	Basic Action	Group
nn	rock	rock		
vb	is	be	be	2
vb	rolling	roll	move	4

Figura 38 — Tabela resultante do programa de análise de texto.

5.3 Módulo #2 — Definição dos tempos das palavras-chave na música

No início deste projecto pretendia-se utilizar sistemas de *speech recognition* para detectar o tempo em que cada palavra surge na música — ao analisar o ficheiro de áudio da música ou o sinal do microfone em *real-time* —, para que o sistema criasse as imagens certas na altura certa da música. Ao experimentar alguns sistemas de *speech recognition* — nomeadamente, a *Cloud Speech-to-Text*⁶³, desenvolvida pela Google, e a *Speech to Text for Java/Processing*⁶⁴, desenvolvida por Florian Schulz —, percebeu-se que existiam diversas limitações que invalidaram a sua utilização neste trabalho. A *Cloud Speech-to-Text* trata-se de uma API desenvolvida pela Google que tem um limite de 60 minutos de análise de áudio em utilização gratuita, após os quais se começa a pagar por cada parcela de 15 segundos de áudio analisado. Concluiu-se, portanto, que o desenvolvimento desta componente do projecto poderia tornar-se insustentável, devido à quantidade de testes necessária. O sistema desenvolvido por Florian Schulz em *Java/Processing* tinha algumas limitações e falhas que poderiam comprometer todo o sistema desenvolvido neste projecto. Em ambos os casos, concluiu-se que os sistemas não estavam preparados para discurso cantado em vez de falado, devido a mudanças no tom e a eventuais prolongamentos de palavras a fim de encaixar na melodia ou ritmo da música, o que torna esta opção ainda mais inviável.

Assim, foi necessário encontrar uma forma de o sistema ter acesso ao tempo em que cada palavra surge na música e optou-se, mais uma vez, por uma via semi-automática. Criou-se, portanto, um pequeno programa em *Processing* que reproduz a canção enquanto conta o tempo em segundos que passou desde o início e dispõe o tempo e cada uma das palavras-chave da letra (substantivos e verbos) no ecrã. A partir da tabela obtida no módulo #1, é possível saber qual é a ordem pela qual estas palavras surgem na música, pelo que o utilizador só tem de acompanhar a letra e premir **ENTER** no preciso momento em que a palavra disposta no ecrã é cantada. À tabela do módulo #1 é adicionada uma coluna onde é guardado o instante em segundos em que cada palavra surge na letra. Desta forma, todas as palavras ficam com um tempo atribuído e é possível, no módulo #4, associar, no tempo certo, as ilustrações e animações a cada substantivo e verbo, respectivamente. Aconselha-se o utilizador a ter a letra presente durante a utilização deste módulo, para ter noção da localização de cada palavra-chave na letra.

Utilizando o exemplo da descrição do módulo #1, *The rock is rolling too fast*, assumindo que o substantivo *rock* surgia no segundo 8 da música, o verbo *is* no segundo 15 e o verbo *rolling* no segundo 19, a tabela resultante seria a da Figura 39. Com esta tabela, será possível, no módulo #4, a ilustração do substantivo *rock* e as animações dos verbos *is* e *rolling* ocorrerem nos instantes em que as palavras são cantadas.

⁶³ Google Cloud. <https://cloud.google.com/speech-to-text/> (09.02.2019)

⁶⁴ Speech Recognition for Java/Processing. <http://florianschulz.info/stt/> (09.02.2019)

POS	Original Word	Word	Basic Action	Group	Time
nn	rock	rock			8
vb	is	be	be	2	15
vb	rolling	roll	move	4	19

Figura 39 — Tabela resultante do programa de definição dos tempos para as palavras-chave.

5.4 Experiências preliminares de ilustração generativa

Nesta secção apresentam-se experiências realizadas numa fase inicial do desenvolvimento do projecto que deram origem a decisões importantes para a estrutura do módulo #3 do sistema.

Foi utilizado o *Processing* para analisar uma imagem que servisse de base para a construção de uma nova imagem estilizada (Fig. 40). Para tal, foi calculada a diferença de cor entre píxeis adjacentes. Se a diferença calculada ultrapassa, em todos os canais de cor, um limite previamente estabelecido (no caso, 30), o píxel assume uma cor diferente (Fig. 41).



Figura 40 — Imagem seleccionada que serviu de base para a ilustração.



Figura 41 — Ilustração gerada.

A fim de evitar o aspecto pixelizado do desenho, experimentou-se desenhar círculos com a posição do píxel como centro (Fig. 42). Os contornos da ilustração tornaram-se mais suaves.

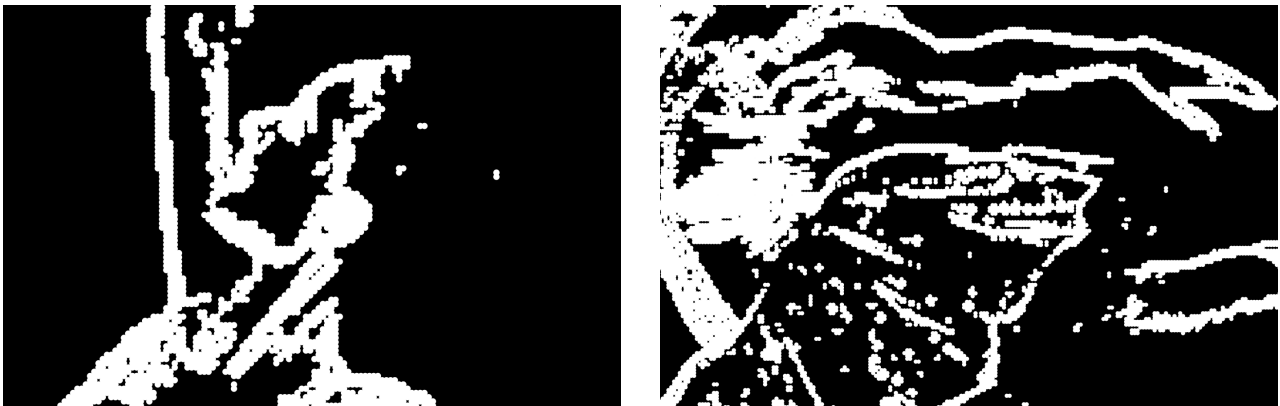


Figura 42 — Experiência com lugar nos píxeis.

Foram feitas algumas tentativas no sentido de ligar os píxeis de modo a desenhar um contorno contínuo, em vez de pontos soltos (Fig. 43). No entanto, não foram obtidos resultados que respondessem ao que se pretendia a nível visual: uma ilustração com um aspecto gráfico interessante, mas que fosse legível.

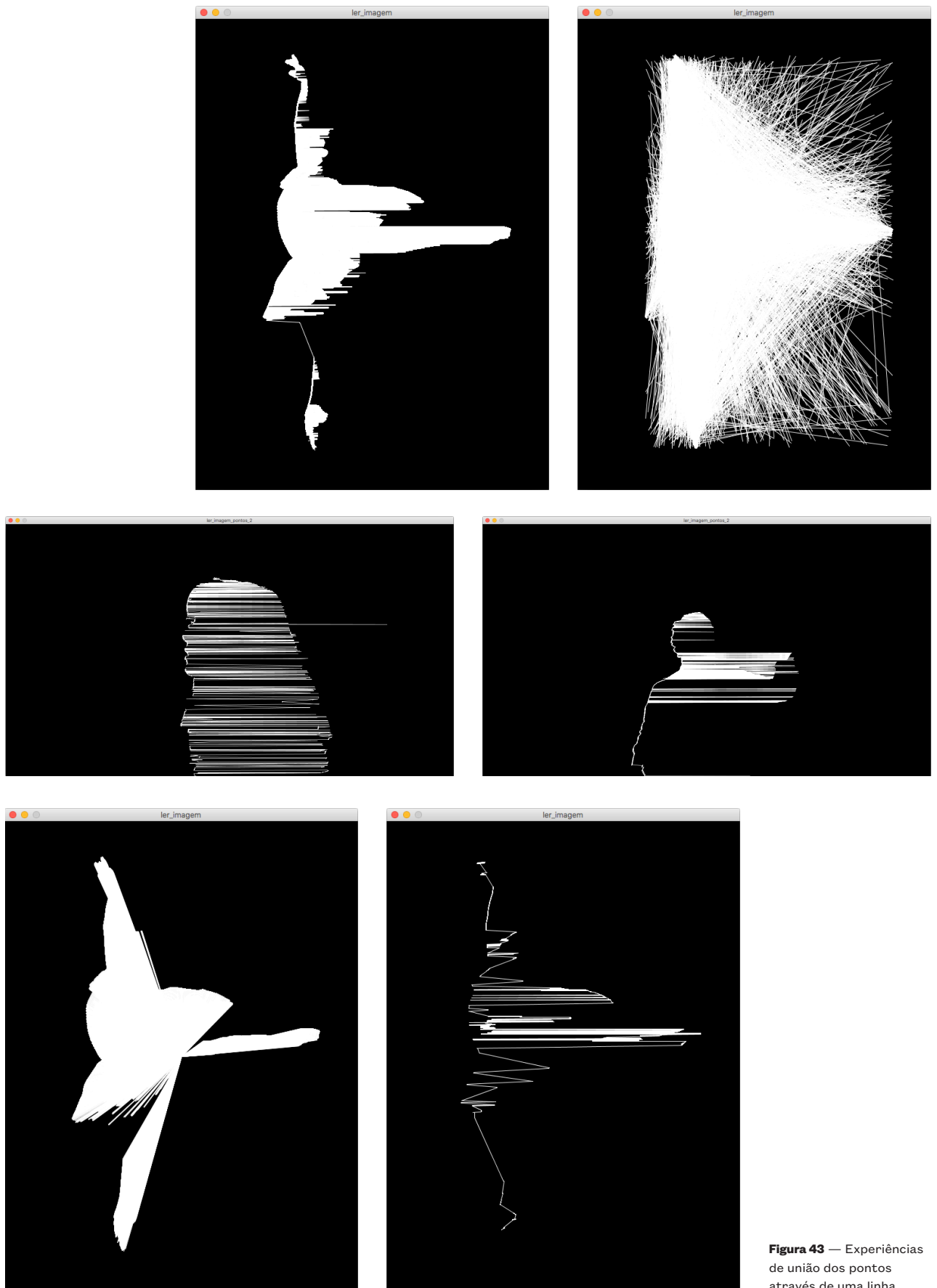


Figura 43 — Experiências de união dos pontos através de uma linha.

Esta abordagem apresentou diversas limitações: **i)** o tempo de processamento era demasiado elevado, **ii)** apenas era legível em imagens cujo elemento principal estivesse bem destacado relativamente ao fundo e **iii)** se o fundo fosse demasiado complexo, além de demorar ainda mais tempo a gerar, o resultado final ficava demasiado confuso e ilegível (Fig. 44).



Figura 44 — Exemplos das duas abordagens numa imagem com fundo complexo.

Assim sendo, houve necessidade de explorar alternativas de ilustração generativa que funcionassem de modo mais eficaz, permitindo uma geração de ilustrações que fosse dinâmica, não só na criação da imagem, mas também na transição entre elas. Por consequência, foi tomada a decisão de dividir o sistema em módulos, para que corresse separadamente e o tempo de processamento na fase de filtração da imagem deixasse de ser um problema. Assim, optou-se por utilizar o *Canny Edge Detector* do *OpenCV Processing*, a fim de detectar de forma eficaz os contornos de uma imagem. Esta ferramenta tem um baixo tempo de processamento e é de simples utilização.

5.5 Módulo #3.1 — Recolha de imagens e aplicação do filtro Canny Edge Detector

Este módulo é responsável pela preparação das imagens que servem de base para as ilustrações criadas no módulo #4. Para que tal aconteça, é necessário recolher ou criar imagens e simplificá-las de modo a que representem uma base para uma ilustração.

Tendo já definido a tabela com as palavras e as relativas POS, é possível identificar e, posteriormente, pesquisar todos os substantivos numa base de dados de imagens. De momento, é pedido ao utilizador que pesquise as imagens manualmente numa plataforma de imagens, a fim de lhe conceder a possibilidade de escolha das imagens e de lhe atribuir alguma liberdade criativa. O carácter humano, pessoal e criativo seria, provavelmente, inexistente num sistema completamente automático. Assim, o utilizador pode, também, manipular as imagens escolhidas ou criar as suas próprias imagens, o que poderá fazer com que as ilustrações resultantes se tornem mais legíveis. Conceitos mais abstractos ou pouco visuais, como, por exemplo, *wind*, são difíceis de encontrar visualmente representados em fotografias, pelo que a melhor alternativa poderá ser criar uma ilustração, mesmo que não literal. No entanto, pretende-se explorar, alternativas automáticas para a pesquisa das imagens, por exemplo, a API da plataforma *Unsplash*.

O utilizador deverá associar quatro imagens diferentes a cada substantivo, para que existam alternativas, no caso de as palavras se repetirem ao longo da música. Isto confere também a possibilidade de haver versões diferentes para a mesma canção, pois as imagens são sempre escolhidas aleatoriamente entre as quatro que representam o mesmo substantivo. Estas imagens podem ser criadas pelo próprio utilizador ou pesquisadas utilizando uma plataforma de imagens como, por exemplo, o *Flickr* ou o *Unsplash*. De momento, o sistema apenas funciona se existirem precisamente quatro imagens para cada substantivo na pasta de onde o *Processing* recolhe as imagens base, uma quantidade fixa que permite alguma diversidade no resultado final. No entanto, pretende-se que o sistema no futuro funcione com qualquer número de imagens, podendo esse número ser diferente em substantivos diferentes, embora seja na mesma necessário que o programa tenha um mínimo e um máximo definidos para o número de imagens. Ao recolher as imagens, o utilizador deverá, também, atribuir um nome ao ficheiro que siga um formato específico que permita ao programa saber, pelo nome do ficheiro **i**) qual é a ordem pela qual as imagens devem surgir ao longo da música — ou seja, a ordem pela qual os substantivos surgem na letra —, **ii**) qual é a palavra relativa a cada imagem e **iii**) qual é o índice da imagem dentro de cada substantivo. Para tal, o nome dos ficheiros das imagens deverá ser algo como **00_bird1.jpg**, **00_bird2.jpg** ou **04_sea2.jpg**, em que os primeiros dois algarismos correspondem à posição da imagem relativamente à ordem pela qual as imagens devem surgir, a palavra a seguir ao *underscore* é o substantivo relativo a essa imagem e o algarismo que se segue é o índice da imagem dentro de cada substantivo.

Numa segunda fase, pede-se ao utilizador que corra o programa que aplica o filtro *Canny Edge Detector* do *OpenCV Processing* às imagens recolhidas. Este filtro detecta os contornos das imagens, resultando na transformação da fotografia numa espécie de desenho simplificado. Este programa permite percorrer as imagens todas com este filtro, ajustar os valores e guardar a nova imagem quando estiver do agrado do utilizador. Para cada imagem é possível ajustar os limites inferior e superior do filtro, a fim de ter mais ou menos informação na imagem. O ideal será encontrar os valores que mantêm a legibilidade da imagem com a menor quantidade de informação (píxeis) possível: quanto mais informação for guardada, mais pesado fica o módulo #4, pois a ilustração fica com mais pontos.

Ao experimentar o programa desenvolvido, obtiveram-se algumas conclusões relativas ao funcionamento deste filtro. Identificou-se que, nesta fase, já se começa a perceber se as imagens seleccionadas pelo utilizador são adequadas ou se precisam de ser substituídas por outras por falta de legibilidade no fim da aplicação do filtro. Abaixo, demonstra-se alguns casos problemáticos resultantes da experimentação do filtro em imagens recolhidas na plataforma de imagens *Unsplash*:

- i) Ao observar a Figura 45, a fotografia original, conseguimos perceber perfeitamente que aquele círculo é o sol e é o motivo principal. No entanto, quando aplicado o filtro, o sol passa apenas a ser um círculo e o motivo principal passa a ser as plantas. Há, portanto, demasiado ruído na fotografia, e falta alguma coisa ao sol para realmente parecer um sol.
- ii) Na Figura 46, percebe-se bem, na fotografia original, que o motivo se trata de balões de ar quente, no entanto, depois da aplicação do filtro, alguns destes balões tornam-se pouco legíveis, problema esse que poderá piorar no módulo #4. Assim, uma opção seria, por exemplo, o utilizador recorrer a um programa de edição de imagens para isolar apenas um dos balões e aumentá-lo, de modo a que o motivo se torne claro.
- iii) Na fotografia original da Figura 47, o motivo principal é o pássaro no ramo de uma árvore. Após a aplicação do filtro, continua a perceber-se o que é, no entanto, o ramo constitui ruído que, no módulo #4, poderá vir a retirar a legibilidade do motivo principal: o pássaro. Assim, uma opção seria, mais uma vez, recorrer a um programa de edição de imagens para isolar o pássaro ou apagar grande parte do ramo.
- iv) Ao observar a Figura 48, na fotografia original nota-se bem a diferença entre o fundo e o motivo, no entanto, com a aplicação do filtro, o motivo mistura-se com o fundo e a imagem torna-se confusa, com muito ruído, demonstrando que tem demasiada informação no fundo.

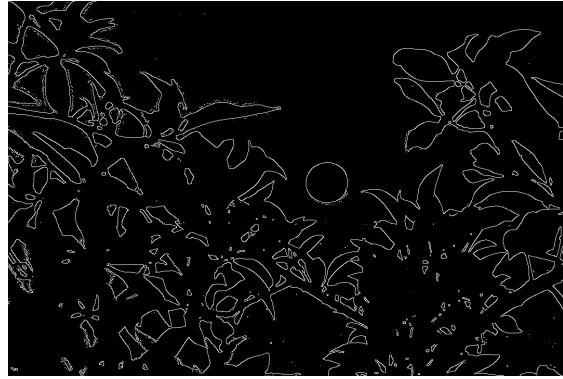


Figura 45 — À esquerda: imagem recolhida do *Unsplash* ao pesquisar a palavra *sun*. À direita: resultado da aplicação do filtro.

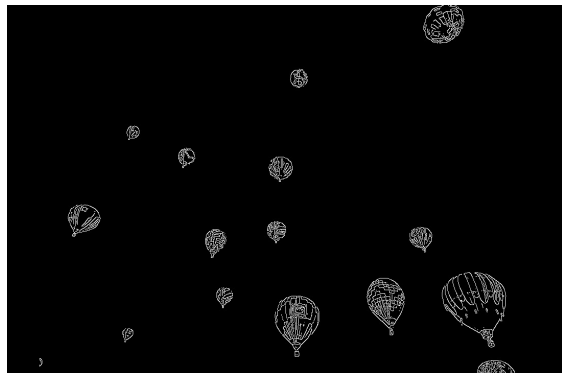


Figura 46 — À esquerda: imagem recolhida do *Unsplash* ao pesquisar a palavra *balloon*. À direita: resultado da aplicação do filtro.



Figura 47 — À esquerda: imagem recolhida do *Unsplash* ao pesquisar a palavra *bird*. À direita: resultado da aplicação do filtro.



Figura 48 — À esquerda: imagem recolhida do *Unsplash* ao pesquisar a palavra *book*. À direita: resultado da aplicação do filtro.

A partir destes exemplos, percebe-se que algumas imagens poderão parecer legíveis no momento em que são recolhidas, no entanto, após a aplicação deste filtro podem perder legibilidade. Este problema que intensifica-se mais ainda no módulo #4, no qual as imagens entram em mais um nível de simplificação (ilustração por pontos). Este tipo de experimentação deverá, portanto, estender-se ao módulo #4, pois algumas imagens poderão permanecer legíveis após a aplicação deste filtro, mas deixar de o ser após a simplificação que acontece no módulo #4. No fim desta fase de experimentação, concluiu-se que as imagens seleccionadas devem ter o mínimo de ruído possível, um fundo o mais homogéneo possível, com o mínimo de motivos possível, a fim de o motivo principal ser rapidamente reconhecido na ilustração resultante do módulo #4. Este é um dos motivos que fortalece a alternativa de o utilizador criar as suas próprias imagens ou manipular as imagens que recolhe.

5.6 Módulo #3.2 — Aprimoramento das imagens base

Nesta secção demonstra-se algumas possibilidades de melhoramento das imagens base a fim de obter mais legibilidade no módulo de ilustração, tendo em conta as anotações presentes na secção anterior. Primeiro demonstra-se o resultado do filtro *Canny Edge Detector* em fotografias editadas com o fim de ter o menor ruído possível e depois demonstra-se a possibilidade de o utilizador criar as suas próprias imagens a partir de ilustrações.

Edição de fotografias

Utilizando alguns dos exemplos acima referidos, considere-se a fotografia original da Figura 49 obtida a partir da plataforma *Unsplash*.

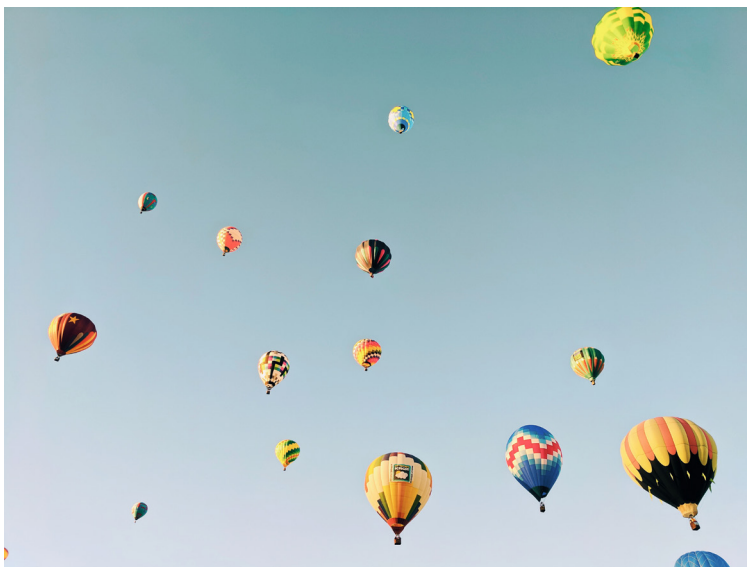


Figura 49 — Imagem recolhida no *Unsplash* ao pesquisar a palavra *balloon*.

Uma simples edição permite seleccionar apenas um dos balões e torná-lo no motivo principal da fotografia. Para tal, seleccionou-se o balão maior e mais direito e eliminou-se o resto da fotografia (Fig. 50). Desta forma, a imagem após a aplicação do filtro torna-se muito mais legível (Fig. 51).



Figura 50 — Imagem após edição num programa de edição de imagens.

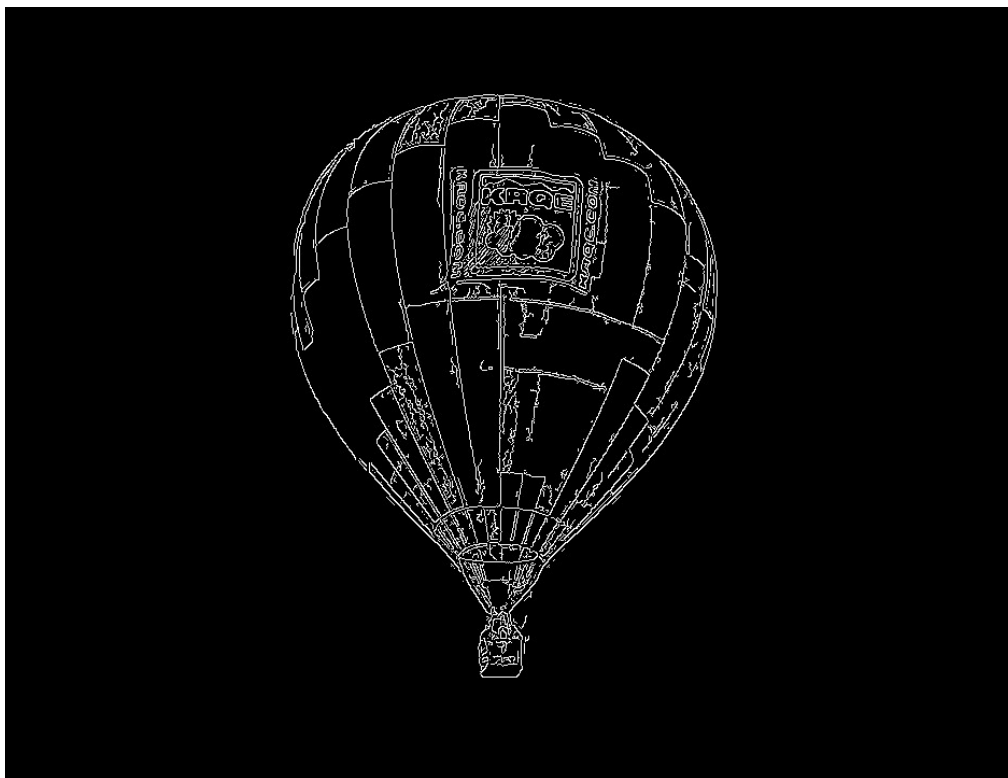


Figura 51 — Imagem resultante da aplicação do filtro.

Considere-se outro exemplo mencionado na secção anterior, presente na Figura 52.



Figura 52 — Imagem recolhida no *Unsplash* ao pesquisar a palavra *book*.

Ao tornar o fundo da fotografia mais homogéneo ou até eliminando-o, deixando apenas o livro, reposicionado para centrar mais na imagem, a fotografia torna-se mais simples (Fig. 53). Pode-se ainda uniformizar mais o interior do livro, também. O resultado após a aplicação do filtro *Canny Edge Detector* pode ver-se na Figura 54.



Figura 53 — Imagem após edição num programa de edição de imagens.

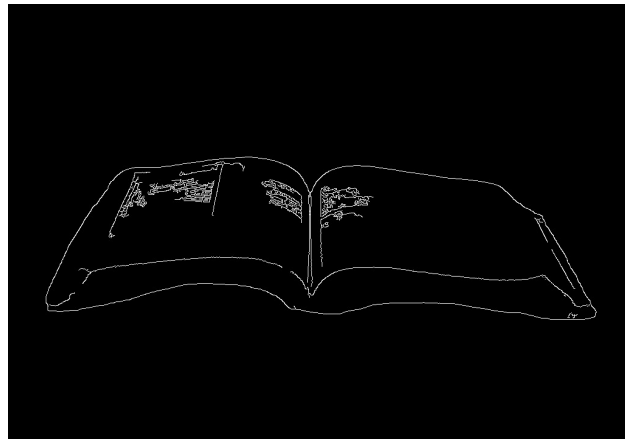


Figura 54 — Imagem resultante da aplicação do filtro.

Como se pode observar nos dois exemplos acima, a pré-edição das fotografias que servem de base para o módulo #4 pode fazer uma diferença significativa na legibilidade final da ilustração a partir de pontos. Este tipo de edição poderá também reduzir o número de píxeis que o filtro *Canny Edge Detector* devolve, melhorando consideravelmente a performance do programa final, além da legibilidade da ilustração.

Criação das imagens a partir de ilustrações

O utilizador poderá também criar as suas próprias imagens, como, por exemplo, captando as suas próprias fotografias, podendo, assim, ter um maior controlo no isolamento do motivo principal ou na posição do motivo que o torna mais reconhecível, ou criando as suas próprias ilustrações, como se demonstra de seguida.

Para experimentar esta abordagem, criou-se quatro ilustrações para cada substantivo presente na letra da canção seleccionada, tal como se fez na pesquisa e recolha de imagens do *Unsplash*. Na Figura 55 exhibe-se alguns exemplos das ilustrações criadas.



Figura 55 — Algumas ilustrações criadas com base da letra da música *Take Me From Here*.

Estas ilustrações foram criadas com vista a serem o mais simplificadas possível, de modo a serem legíveis e terem pouco ruído. No entanto, neste caso não faz muito sentido aplicar o filtro *Canny Edge Detector*, pois as próprias ilustrações já são contornos. O resultado da aplicação do filtro é apresentado na Figura 56.



Figura 56 — Resultado da aplicação do filtro *Canny Edge Detector* numa ilustração.

Apesar de a imagem se manter legível, tem informação em duplicado, pois está a extrair o contorno de um contorno. Assim, faz mais sentido utilizar a própria ilustração como base do módulo #4. Porém, as ilustrações foram criadas com linhas muito grossas, o que faz com que o número de píxeis resultantes seja maior do que aquele que é necessário. Desta forma, aprimorou-se as ilustrações utilizando uma caneta de ponta mais fina e, após edição para ficar com as cores certas — fundo preto ($\text{RGB}(0, 0, 0)$) e linhas brancas ($\text{RGB}(255, 255, 255)$), tal como no *Canny Edge Detector* — o resultado é o que se apresenta na Figura 57.



Figura 57 — Resultado da aplicação do filtro *Canny Edge Detector* numa ilustração.

Com estas ilustrações simplificadas, o resultado torna-se consideravelmente mais legível do que utilizando fotografias, pois, mesmo editando as fotografias, há sempre informação não desejável na imagem base que o filtro *Canny Edge Detector* detecta como contorno.

5.7 Módulo #4 — Ilustração e animação com base nas imagens filtradas

Este módulo é responsável pela ilustração de cada substantivo presente na letra da música e pela sua animação de acordo com os verbos que surgem posteriormente. A ilustração é composta por pontos que se movem de modo a criar as animações ou manipular a imagem geral de acordo com atributos sonoros, como será explorado no módulo #5. Se surgir um substantivo na letra da música, os pontos deslocam-se para as suas novas posições de modo a desenhar a imagem que o representa. Apesar de essa imagem se manter até surgir o próximo substantivo, existem animações nos pontos de acordo com os verbos que vão surgindo na letra.

Para criar as ilustrações, tem-se por base as imagens criadas no módulo #3. O programa percorre a imagem base correspondente ao substantivo actual (Fig. 58) e, sempre que passa por um pixel branco, desenha um ponto nessa posição (Fig. 59). Para trocar de imagem, os pontos só têm de se deslocar de uma posição para outra que desenhe a imagem seguinte.



Figura 58 — Imagem base.



Figura 59 — Ilustração com pontos.

Um dos problemas encontrados ao longo do desenvolvimento deste módulo, foi o facto de cada imagem ter um número diferente de pontos, dependendo da sua complexidade. Isto significa que é necessário acrescentar ou eliminar pontos em cada transição. Para que estas acções não retirassem fluidez às transições, encontrou-se a solução que se descreve de seguida para o funcionamento do programa sempre surge uma transição:

- i) No caso de a nova imagem ter mais pontos do que a anterior, acrescentar os novos pontos nas mesmas posições que alguns dos pontos da imagem anterior escolhidos aleatoriamente e fazê-los deslocarem-se para as suas posições da nova imagem.
- ii) No caso de a nova imagem ter menos pontos do que a anterior, fazer os pontos que estão a mais deslocarem-se para as mesmas posições de alguns dos pontos da imagem nova escolhidos aleatoriamente, eliminando-os no fim.

Desta forma, é possível ter uma transição fluida sem cortes repentinos ao acrescentar ou eliminar pontos.

O processo de desenvolvimento das animações implicou uma revisão no código de modo a existir uma classe **Point** que guardasse algumas variáveis importantes de cada ponto, como, por exemplo, a posição actual e a posição objectivo, para realizar as transições, ou, para algumas animações, a sua forma, tamanho, cor ou transparência. Assim, é possível não só controlar o sistema de partículas, mas também cada ponto individualmente. As funções responsáveis pelas transições e animações estão também inseridas nesta classe. Esta revisão no código facilitou consideravelmente o desenvolvimento posterior do programa e melhorou o seu processamento.

Numa primeira fase, tentou-se desenvolver as animações provisórias definidas no módulo #1. No entanto, ao longo do processo, encontrou-se diversos problemas e obtiveram-se conclusões que obrigaram a uma redefinição das animações.

No início, definiu-se uma animação para o grupo #1 (*smell, see, hear, feel*) que deslocasse os pontos para um círculo, uma representação abstracta escolhida para representar os cinco sentidos. Os cinco sentidos são métodos de percepção do meio ambiente que nos rodeia a partir de determinadas partes do nosso corpo especificamente dedicadas a essa função. Todas estas partes estão ligadas ao mesmo ponto central, o cérebro, que descodifica a informação que lhe chega a fim de a reconhecer. Tendo estes aspectos como inspiração, considerou-se que o círculo seria o símbolo abstracto mais próximo deste conceito, pois trata-se da representação de todos os pontos que se encontram à mesma distância de um ponto central. Além disso, grande parte das representações esquemáticas dos cinco sentidos consistem nos sentidos dispostos num círculo. No entanto, quando se colocou esta ideia em prática, a animação parecia uma transição igual à que acontece quando surge um novo substantivo, ou seja, o círculo poderia confundir-se com uma ilustração (Fig. 60).

Concluiu-se, então, que as animações definidas para os verbos não podem alterar as posições principais dos pontos, não só para não parecer uma transição, mas também para não retirar a legibilidade da imagem. Assim, de modo a manter a representação do círculo, colocou-se cada um dos pontos a rodar em torno de um pequeno círculo com centro na sua posição base.

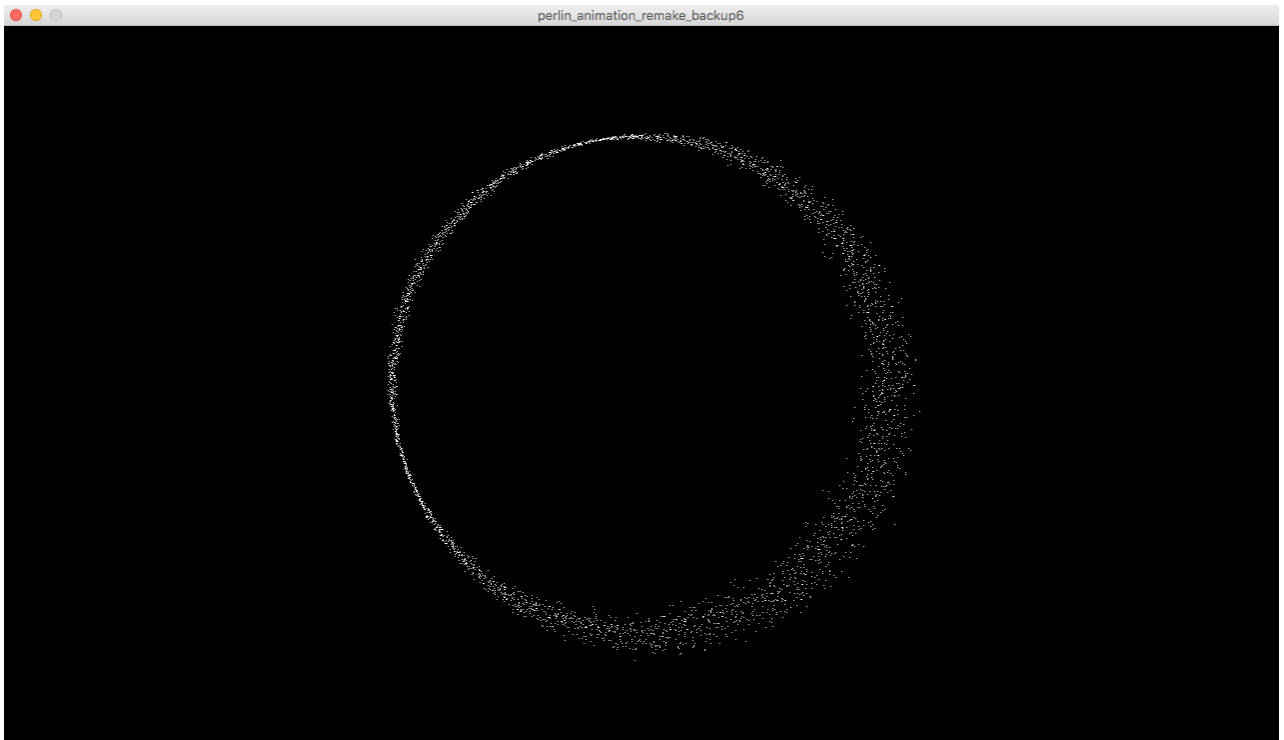


Figura 60 — Animação definida inicialmente para o grupo #1.

Para o grupo #2 (*conclude, think, be*), tentou-se atribuir um *glow* ou *blur* a cada um dos pontos, de modo a representar a ideia abstracta de existência ou pensamento. No entanto, todos os métodos testados para criar um *glow* ou um *blur* diminuíram consideravelmente a velocidade de processamento do programa, tornando-o lento durante a animação. Assim, atribuiu-se uma nova animação que consiste em fazer os pontos piscar aleatoriamente. Esta representação teve como inspiração interpretações subjectivas do que é, a nível visual, o pensamento. Poderá ser uma representação visual i) das sinapses que ocorrem nos neurónios com o objectivo de transmitir informação para cada parte do cérebro e produzir uma percepção ou pensamento, ou ii) das estrelas no céu, que sempre foram objecto de estudo de muitos cientistas e filósofos, a fim de obter conclusões relativamente à nossa presença no universo ou à nossa existência. Para tal, trabalhou-se com a transparência de cada ponto individualmente, o que seria algo mais complexo, tanto no desenvolvimento, como no processamento, se não se tivesse criado a classe **Point**.

A animação inicialmente definida para o grupo #3 (*speak*) — colocar os pontos a vibrar mais intensamente — foi atribuída ao grupo #4 (*move, transport*), pois o conceito da vibração do som poderia ser confundido com a modelação da imagem de acordo com o sinal sonoro efectuada no módulo seguinte. Assim, esta vibração nos pontos representa movimento e não vibração sonora. Para a animação da acção básica *speak* definiu-se uma representação alternativa: colocou-se os pontos a aumentar e diminuir aleatoriamente de tamanho, como se fossem pessoas a falar, representando o burburinho de uma multidão. Também nesta animação o desenvolvimento

da classe `Point` foi importante, pois o tamanho de cada ponto é afectado individualmente.

A animação do grupo #4 (*move, transport*), como já foi mencionado, acabou por ser a vibração dos pontos, criada a partir de um movimento rápido dos pontos em direcções aleatórias, a fim de transmitir a ideia de movimento.

Na animação do grupo #5 (*ingest, have*), surgiu o mesmo problema da primeira animação do grupo #1: a legibilidade da imagem ficou comprometida ao fazer um deslocamento de todos os pontos para o centro (Fig. 61).

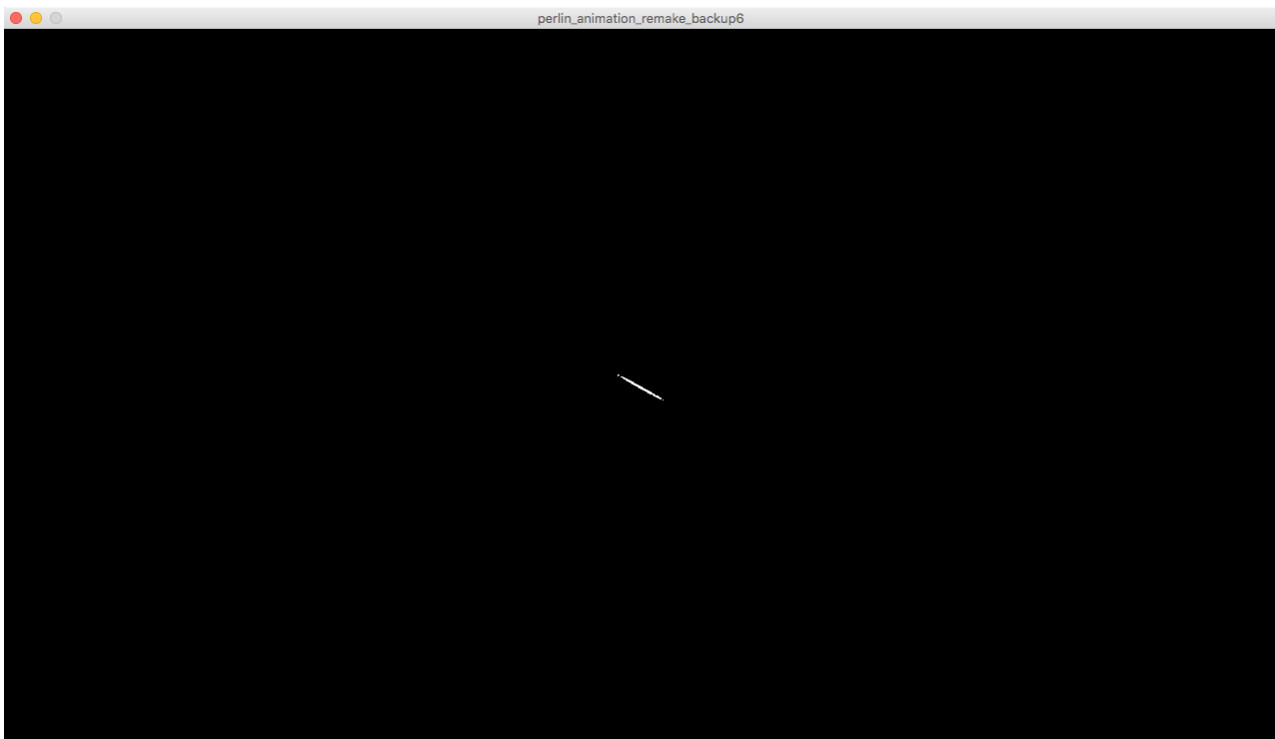


Figura 61 — Animação definida inicialmente para o grupo #5.

A fim de manter a ideia base, a animação consiste no deslocamento dos pontos em direcção ao centro, sem nunca lá chegar, resultando num efeito de diminuição da escala da imagem. Como esta animação estava menos interessante do que as restantes, pois parecia apenas uma simples alteração de escala da imagem, optou-se por se colocar velocidades diferentes nos pontos enquanto se deslocam para o centro, criando a ilusão de os pontos se estarem a deslocar num espaço 3D.

Aplicou-se o mesmo efeito, mas de modo inverso para a animação do grupo #6 (*expel, propel*), com deslocação dos pontos para fora, na direcção oposta à do centro, resultando num efeito de aumento da escala da imagem. Também aqui, e pelos mesmos motivos da animação anterior, se colocou velocidades diferentes nos pontos enquanto se deslocam para fora.

As animações seleccionadas, quando aplicadas, atribuem mais dinamismo e tornam o resultado final mais interessante. No entanto,

por vezes, a animação não se relaciona muito com o significado do verbo, não só por causa da limitação já referida da função `getDistance()` da biblioteca *RiTa* do *Processing*, mas também por todos os verbos serem mapeados apenas para seis grupos. Por exemplo, os verbos *push* e *come* são associados pela função `getDistance()` à acção básica *see*, o que significa um mapeamento do verbo ao grupo #1, que corresponde aos cinco sentidos. Pretende-se, no futuro, explorar mais animações, nomeadamente, uma para cada uma das quatorze acções básicas. Esta exploração poderá trazer ainda mais dinamismo ao resultado final e uma aproximação da animação ao significado do verbo correspondente.

5.8 Módulo #5 — Mapeamento de atributos sonoros com variáveis visuais

Este módulo consiste na análise de diferentes atributos do sinal sonoro de uma música e no seu mapeamento com algumas variáveis visuais da ilustração. Estudou-se diversas possibilidades de mapeamento: **i)** directo, **ii)** através de emoções ou **iii)** através de outros tipos de intermediários. O resultado final ainda está muito aberto a exploração futura, não só por falta de conhecimento técnico, mas também devido a impossibilidade tecnológica. O objectivo inicial desta componente consistia em representar visualmente as emoções contidas numa música. No entanto, as tecnologias actualmente existentes para detecção de atributos sonoros não têm precisão suficiente para detecção de emoções de forma satisfatória. Além disso, explorar essa área seria algo demasiado complexo no contexto desta dissertação. De seguida, apresentam-se algumas referências desta área de investigação que pretende reconhecer as emoções contidas numa música a partir do seu sinal sonoro.

A maioria dos estudos analisados sobre o reconhecimento de emoções em música utilizam o modelo de Thayer (Thayer, 1989) como base. Este é um modelo bidimensional em que o eixo dos XX se refere ao Stress (*Valence*) e o eixo dos YY à Energia (*Arousal*). A *Valence* altera-se entre ansioso (negativo) e alegre (positivo) e o *Arousal* varia entre calmo (baixo) e energético (alto). Assim, os quatro quadrantes resultantes são (Fig. 62):

- i) Energético-Negativo: Ansioso
- ii) Energético-Positivo: Exuberante
- iii) Calmo-Negativo: Depressivo
- iv) Calmo-Positivo: Pacífico/Contente

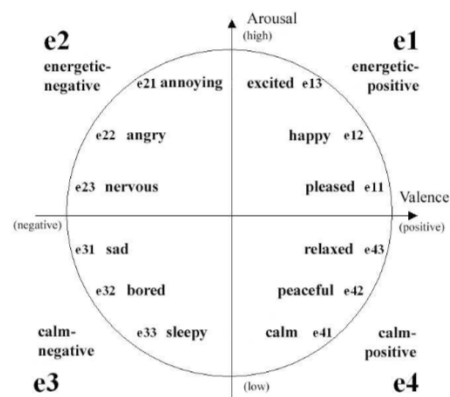


Figura 62 — Diagrama do modelo de Thayer. Screenshot do artigo de Grekow e Ras (1970).

Os estudos analisados dividem-se em duas categorias: análise de ficheiros MIDI e análise do sinal sonoro. O que interessa mais para a presente dissertação é a análise do sinal sonoro, uma vez que a análise de ficheiros MIDI é limitadora a muitos níveis. Poderia ser interessante estudar esta via, se a música em questão se tratasse apenas de piano e voz, por exemplo. Nesse caso, a análise do ficheiro MIDI com intuito de adquirir a emoção do piano, tanto num concerto (*real-time*) como numa gravação de estúdio, poderia ter resultados visuais interessantes e bastante fiéis ao som. Além disso, poderia associar-se visualizações diferentes para a voz e para o piano. Embora para a presente implementação esta via não seja muito relevante, pretende-se explorá-la no futuro, a fim de enriquecer a componente dos efeitos visuais dos concertos.

Assim, os estudos que foram analisados mais a fundo são descritos de seguida.

Music Emotion Recognition, de Yi-Hsuan Yang e Homer H. Chen

Este estudo (Yang & Chen, 2011) foi realizado em 2011 e descreve exhaustivamente todo o processo de análise de uma música para fins de reconhecimento da sua emoção. Começa por explicar a importância desta área, os problemas que ainda existem — como, por exemplo, a ambiguidade na descrição de uma emoção, a subjectividade da percepção de uma emoção e o salto que existe entre os atributos sonoros de baixo e médio nível e a percepção humana de alto nível —, faz uma visão global da descrição e reconhecimento de emoções em música e, por fim, chega aos atributos sonoros e musicais analisados e vagamente associados aos dois eixos do modelo de Thayer. Para este projecto foi dado um maior enfoque a este último capítulo, pois foi necessário conhecer os atributos sonoros que influenciam a percepção da emoção da música e como influenciam essa percepção, para que se pudesse utilizá-los de forma correcta. O resto do livro de Yang e Chen descreve o algoritmo e o sistema de identificação de emoções em músicas desenvolvido pelos autores e, como esta componente não entra no foco deste trabalho, não foi estudado a fundo.

No início do estudo, os autores criaram um *dataset* de canções que dividiram pelos quatro quadrantes do modelo de Thayer de acordo com a emoção percebida. Para efeitos de demonstração no seu livro, seleccionaram aleatoriamente quatro canções pertencentes a este *dataset*, cada uma representante de um dos quadrantes:

- i) *Smells Like Teen Spirit* — Nirvana:
Quadrante energético-negativo (ansioso)
- ii) *Are We The Waiting* — Green Day:
Quadrante energético-positivo (exuberante)
- iii) *Mad World* — Gary Jules:
Quadrante calmo-negativo (depressivo)
- iv) *White Christmas* — Lisa Ono:
Quadrante calmo-positivo (pacífico/contente)

Desta forma, ao exibir os valores obtidos na análise de cada música, Yang e Chen demonstraram as suas conclusões relativamente a cada atributo e à sua ligação com a *Valence* ou o *Arousal*. O estudo divide os atributos em atributos de energia, de ritmo, temporais, espectrais e harmónicos. Os autores concluem que **i)** na categoria dos atributos de energia, o atributos (por exemplo, *loudness*) estão mais relacionados com o *Arousal*, **ii)** nos atributos de ritmo, a maioria deles também estão mais relacionados com o *Arousal*, apesar de a fluência do ritmo poder ser associada à *Valence*, **iii)** dentro dos atributos temporais, conclui-se que a *zero-crossing rate* está associada ao *Arousal*, enquanto o desvio padrão da *zero-crossing rate* poderá estar associada à *Valence*, **iv)** nos atributos espectrais, mencionam-se diversos atributos, sem mencionar a sua relação com os dois eixos do modelo de Thayer, no entanto, menciona-se que a *spectral flatness* e o *spectral crest factor* podem estar relacionados com a *tonalness* e, portanto, com a *Valence*, que o *spectral centroid* e o desvio-padrão do *spectral rolloff* estão associados ao *Arousal*, e que a *Roughness* está associada à *Valence* da música, **v)** e, por fim, nos atributos harmónicos, conclui-se que a média do *pitch* está associado ao *Arousal* e o seu desvio padrão à *Valence*, o modo musical (maior ou menor) também está fortemente associado à *Valence*, no entanto, ainda é muito difícil obter o modo musical num sinal sonoro. Todas estas conclusões são muito vagas, no entanto, percebeu-se que é muito mais fácil e há muito mais atributos para analisar o *Arousal* de uma música do que a sua *Valence*.

Automatic Mood Detection from Acoustic Music Data,
de Dan Liu, Lie Lu e Hong-Jiang Zhang

Este estudo (Liu *et al.*, 2003) é de 2003, no entanto, os mesmos autores continuaram o estudo e em 2006 publicaram um novo artigo mais desenvolvido e com uma componente de *tracking* do *mood* ao longo da música, além da sua deteção: *Automatic Mood Detection and Tracking of Music Audio Signals*. Ao ler os artigos, percebeu-se que a componente de deteção do *mood* da música, portanto, a mais relevante para o presente trabalho, manteve-se semelhante nos dois. É explicitada no segundo artigo a diferença entre emoção e *mood*: emoção, habitualmente, refere-se a uma experiência curta mas forte, enquanto *mood* é uma experiência mais longa e menos forte. Os autores defendem que, neste caso, o significado de *mood* se adequa mais.

Ainda na introdução é apresentada uma justificação para a necessidade de lidar com sinal acústico em vez de MIDI. Apesar de os estudos que utilizam MIDI para reconhecer emoções obterem resultados impressionantes, a maioria da música existente no mundo real não está em MIDI e, tanto quanto foi possível verificar, ainda não existem sistemas de transcrição que consigam traduzir sinais acústicos em representações simbólicas como o MIDI. Assim, continua a ser importante conseguir reconhecer emoções em músicas a partir do sinal sonoro.

Também neste estudo se utiliza o modelo de Thayer para mapear os atributos do sinal sonoro às emoções. Os atributos dividem-se em atributos

de intensidade, timbre e ritmo, no entanto, nestes dois artigos não foi possível identificar de modo explícito a influência de cada um dos atributos no *Arousal* ou na *Valence*.

Os autores concluem, ainda, que o modo musical é muito difícil de obter a partir de sinais sonoros, por isso, apenas utilizam atributos de intensidade, timbre e ritmo. Os de intensidade referem-se ao *Arousal* e os de timbre e ritmo à *Valence*. Assim, o sistema desenvolvido neste estudo utiliza uma estrutura hierárquica, colocando, numa primeira fase, a intensidade da música a dividir os quatro quadrantes do modelo de Thayer em dois, por ser mais fácil de calcular computacionalmente: a energia dos quadrantes Contente e Depressivo (baixo *Arousal*) é muito menor do que a energia dos quadrantes Exuberante e Ansioso (*Arousal* elevado). Assim, se a energia da música for baixa, irá para um grupo, se for elevada, irá para o outro.

Numa segunda etapa do sistema, o grupo em que a música se encontra será dividido de acordo com outros atributos, a fim de descobrir em qual dos dois quadrantes desse grupo a música se insere (Contente/Depressivo ou Exuberante/Ansioso, dependendo do grupo). Para o timbre, são analisados atributos espectrais, como *spectral centroid*, *spectral rolloff* e *spectral flux*. A conclusão, nesta fase, é de que o quadrante Exuberante contém valores de *spectral centroid* muito mais elevados do que o quadrante Depressivo, pois as músicas com mais *Arousal* tendem a ter um espectro maior de frequências altas, e as de menor *Arousal* um espectro maior de frequências baixas. No segundo artigo, ainda se refere que o *brightness* do quadrante Exuberante é mais elevado do que o do quadrante Depressivo. Quanto ao ritmo, é, geralmente, analisado com base na força, regularidade e tempo. Uma das conclusões exibidas no artigo é que músicas pertencentes ao quadrante Exuberante tendem a ter um ritmo forte, estável e rápido, enquanto no Depressivo, as músicas são mais lentas e sem padrão rítmico distinto. Em ambos os artigos, tal como na maioria da bibliografia analisada, não há grande desenvolvimento neste tipo de conclusões, pelo que se torna complicado recolher informação útil, principalmente, sobre os atributos que afectam a *Valence*.

Musical Genre Classification of Audio Signals,
de George Tzanetakis e Perry Cook

Este estudo (Tzanetakis e Cook, 2002) de 2002 foi também analisado, no entanto, não contribuiu muito na investigação da influência dos atributos sonoros no reconhecimento da emoção de uma música. Assim, a conclusão pertinente e relevante para a presente dissertação é que também se utilizaram atributos tímbricos (*spectral centroid*, *spectral rolloff*, *spectral flux*, *zero-crossing rate*, *MFCC*, entre outras) e rítmicos para classificar o género musical de um sinal sonoro. Assim, percebe-se que existe um padrão em todos os estudos analisados: todos eles se debruçam nos atributos de intensidade, tímbricos e rítmicos, com especial destaque para a análise espectral.

A Survey of Audio-Based Music Classification and Annotation,
de Zhouyu Fu, Guojun Lu, Kai Ming Ting e Dengsheng Zhang

Este é um estudo (Fu *et al.*, 2011) de 2011 que tem como objectivo resumir um pouco o estado da arte nesta área da *music information retrieval* (MIR) e definir os aspectos essenciais para cada tipo de classificação de música. No caso da presente dissertação, apenas é relevante a secção em que referem os atributos de áudio necessários para classificar música. É apresentada a distribuição dos diferentes atributos pelo baixo, médio e alto nível, não esquecendo o já referido salto entre o nível médio e o nível alto (Fig. 63).

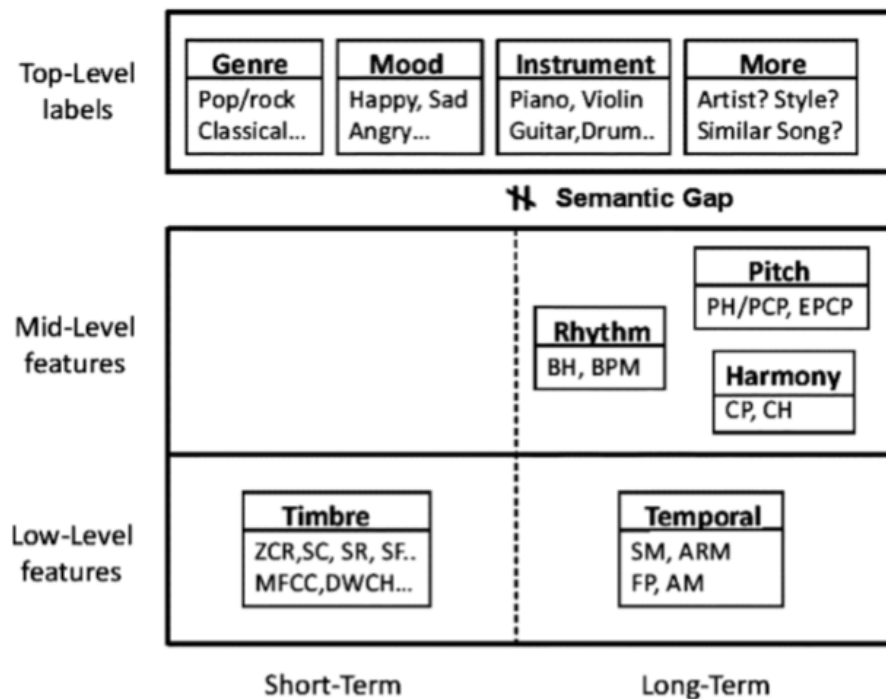


Figura 63 — Distribuição dos diferentes atributos sonoros pelos baixo, médio e alto níveis. Screenshot do artigo de Fu *et al.* (2011).

É importante ter uma noção desta organização dos atributos quando se quer fazer uma análise de um sinal sonoro, pois a distância entre o baixo nível e o alto nível é extremamente elevada. Assim, para a presente dissertação é importante que esses atributos sejam perceptíveis pelo ouvido humano, uma vez que se pretende mapear atributos sonoros com variáveis visuais. A partir deste diagrama, podemos concluir que os atributos de baixo nível deverão afectar variáveis visuais de forma mais indirecta, enquanto e os de nível médio poderão, então, ter um mapeamento mais directo.

Reunião com o Prof. Dr. Rui Pedro Paiva

O Prof. Dr. Rui Pedro Paiva, Professor no Departamento de Engenharia Informática da Universidade de Coimbra, desenvolveu em Outubro de 2012,

juntamente com o seu aluno doutorando Renato Panda, um algoritmo de classificação de *mood* a partir de áudio e continuam ambos a aprofundar conhecimentos na área de classificação de emoções de acordo com atributos sonoros⁶⁵.

Numa reunião via *Skype*, discutiu-se o sucesso dos diversos estudos efectuados ao longo das duas últimas décadas no âmbito do reconhecimento de emoção a partir de um sinal sonoro, nomeadamente do estudo *Music Emotion Recognition* de Yi-Hsuan Yang e Homer H. Chen, que foi analisado em detalhe para este trabalho. Concluiu-se que este problema ainda se encontra longe de estar resolvido, apesar de já existirem diversas soluções válidas e satisfatórias. A previsão da *Valence* no estudo de Yang e Chen (2011) tem uma precisão muito menor do que a do *Arousal*. No projecto de Paiva e Panda, a precisão da previsão do *Valence* é consideravelmente maior, apesar de estar ainda longe de ser perfeita. Além disso, Paiva colocou alguns problemas relativamente ao projecto de Yang e Chen: **i**) o número de atributos avaliados é bastante limitado, **ii**) o *dataset* (conjunto de músicas) utilizado é ainda reduzido e **iii**) a concordância dos resultados com as anotações do *dataset* é baixa. Tudo isto poderá diminuir a precisão da avaliação do sistema desenvolvido. Relativamente ao seu projecto com Panda, Paiva refere que se deu “um passo em frente na construção de um *dataset* de qualidade” (constituído por mais de 900 músicas) “e na criação de novas *features* com relevância emocional”, no entanto, é necessário “continuar o trabalho de identificação de *features* importantes. É também importante melhorar alguns dos algoritmos subjacentes”, por exemplo, “detecção de melodia em áudio polifónico, tempo, tonalidade, etc.”. Concluiu-se também que muita da imprecisão presente nas avaliações dos diversos estudos se deve ao factor de erro presente nos algoritmos existentes e disponíveis hoje em dia. No entanto, o principal factor que, provavelmente, diminui a precisão neste tipo de avaliação é a quantidade de variáveis que existem num sinal sonoro que podem influenciar toda a percepção da emoção numa música. Apesar de existirem, de facto, atributos do sinal sonoro que estão geralmente associados à *Valence* e outros ao *Arousal*, estes podem assumir valores que não batem certo com a percepção humana da *Valence* ou *Arousal* numa música. Só quando avaliados em conjunto é que se consegue obter conclusões mais precisas relativamente a estes atributos.

Tendo em conta que prever a emoção presente numa música ainda se revela um desafio complexo, como se pôde constatar nos diversos estudos descritos acima, concluiu-se que seria mais relevante para o presente trabalho explorar alguns atributos referentes a diferentes características do som — intensidade, timbre, ritmo e melodia —, estudar os seus conceitos e seleccionar variáveis visuais que melhor se adequam a cada um desses conceitos. Desta forma, ao realizar um mapeamento entre os atributos sonoros e as variáveis visuais seleccionadas, torna-se possível obter resultados visuais que se aproximem da emoção transmitida pela música. No entanto, nem todos os atributos sonoros são perceptíveis pelo ouvido

humano, pelo que um mapeamento entre estes atributos com variáveis visuais poderia gerar ruído visual ou confusão ao ouvinte. Assim, chegou-se à conclusão de que as variáveis visuais mais evidentes deveriam ser influenciadas pelos atributos sonoros mais perceptíveis. Os atributos sonoros menos perceptíveis podem ser mapeados com variáveis visuais menos evidentes, a fim de ajudar na dinâmica e na transmissão de uma emoção visual.

Como já foi mencionado, no estudo *Music Emotion Recognition* de Yang e Chen (2011) foram escolhidas quatro canções que representam cada um dos quatro quadrantes do modelo de Thayer:

- a) *Smells Like Teen Spirit* — Nirvana (Ansioso):
 - Valence
 + Arousal
- b) *Are We The Waiting* — Green Day (Exuberante):
 + Valence
 + Arousal
- c) *Mad World* — Gary Jules (Depressivo):
 - Valence
 - Arousal
- d) *White Christmas* — Lisa Ono (Pacífico/Contente):
 + Valence
 - Arousal

Durante a implementação do sistema do presente trabalho, optou-se por também utilizar estas canções para analisar na perspectiva de cada um dos atributos sonoros e, utilizando o *Pure Data*, obtiveram-se resultados que foram confrontados com as conclusões descritas no estudo de Yang e Chen (2011). Embora uma amostra de quatro músicas não seja, de todo, suficiente para comprovar as escolhas realizadas ao longo deste projecto, no tempo disponível só foi possível tomar as decisões com base nos resultados obtidos nestas quatro músicas. Quando comparados, os valores obtidos em cada análise eram bastante diferentes, possivelmente porque as ferramentas utilizadas no estudo de Yang e Chen (*PsySound*, *SDT*, *Marsyas*, *MA toolbox*, *MIR toolbox*, *RP extractor* e *MATLAB*) utilizam algoritmos diferentes da que foi utilizada neste trabalho (*Pure Data*), por isso, analisou-se apenas as conclusões relativas à influência de cada atributo sonoro na percepção de *Arousal* ou *Valence* e comparou-se com as retiradas neste trabalho, que serão apresentadas de seguida. Optou-se por reduzir o número de atributos utilizados para os que estão mais consistentes com as conclusões de todos os estudos analisados, os que se revelam bastante claros quando analisados na pequena amostra de quatro músicas e os que se poderiam relacionar melhor com variáveis visuais: amplitude, *spectral centroid* e *spectral flux*. Alguns dos atributos seleccionados são utilizados em mais do que uma variável visual. Pretende-se, no futuro, explorar mais atributos e a sua relação com variáveis visuais, nomeadamente *pitch*, tempo, *onset strength*, *average onset*, *zero-crossing rate*, *spectral rolloff* e *roughness*, entre outros.

De seguida serão apresentados os atributos utilizados e o seu mapeamento com as variáveis visuais e será descrita análise realizada aos atributos não utilizados.

Amplitude

A amplitude corresponde à energia do sinal sonoro.

Utilizando o objecto [env~] do *Pure Data*, que analisa a amplitude do envelope do sinal sonoro em cada *frame*, calculou-se a média da amplitude de cada canção (numa escala de 0 a 100):

- a) 80,64 dB
- b) 88,35 dB
- c) 80,80 dB
- d) 82,00 dB

Analisando estes resultados, chegou-se à conclusão de que os valores obtidos não correspondem aos valores esperados, pois as canções com mais *Arousal*, a) e b), deveriam ter uma média de amplitudes consideravelmente mais elevada do que as canções com menos *Arousal*, c) e d). Estes valores poderão ter sido influenciados pela diferença na masterização dos quatro ficheiros, o que afecta consideravelmente a resposta do objecto [env~] do *Pure Data*. Mesmo com normalização dos volumes no *Audacity*, através da ferramenta *Normalize*, não foi possível obter os valores esperados.

Assim, concluiu-se que, a nível visual, a amplitude deve ter uma representação relativa, e não absoluta, de modo a visualizar a intensidade da música em cada momento relativamente à média do resto da música. Se a amplitude actual for maior do que a média da amplitude, a luminosidade do fundo incrementa um valor fixo, se for menor, decrementa. Desta forma, sempre que a música está mais intensa, o fundo torna-se mais brilhante, sempre que se acalma, o brilho do fundo desce de novo até ao preto. Na Figura 64 e Figura 65, pode ver-se, respectivamente, um *frame* da música em que o brilho está no mínimo e outro em que o brilho está no máximo. A *color hue* será descrita mais à frente.

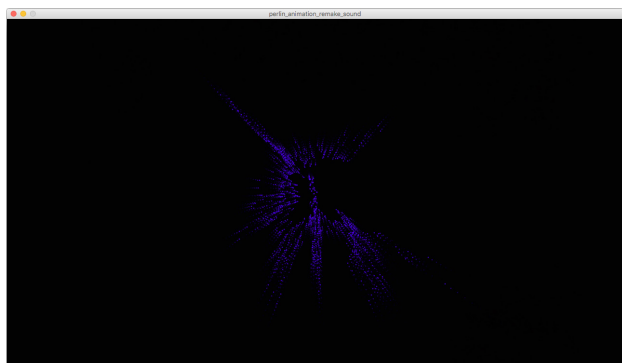


Figura 64 — *Frame* com o nível de brilho no mínimo.



Figura 65 — *Frame* com o nível de brilho no máximo.

Além desta funcionalidade, a amplitude também atribui energia ao *perlin noise* presente no movimento dos pontos. Sempre que a amplitude actual for maior do que a amplitude média, é adicionada energia aos pontos, sempre que for menor, é retirada. Assim, é possível perceber visualmente que quanto maior o movimento dos pontos e mais brilhante o fundo, mais intensa é essa secção da música relativamente ao resto. Conseguir-se, desta forma, ter uma representação visual da variação de intensidade ao longo da música.

Até aqui, a intensidade da música é representada apenas numa perspectiva geral da música, pelo que a intensidade da música em cada instante não é facilmente percebida. Assim, optou-se por utilizar também a amplitude para afectar directamente uma variável visual em cada instante. Para tal, decidiu-se utilizar a amplitude para definir a forma geométrica dos pontos quando estes são desenhados. Escolheram-se quatro formas diferentes, cada uma associada a um nível de amplitude. As formas, por ordem crescente de amplitude, são: **ponto**, **cruz**, **círculo** e **rectângulo**.

Numa primeira experiência, definiu-se quatro intervalos de amplitude e associou-se uma das formas, por esta ordem, a cada um. Por exemplo, se a amplitude estivesse entre 0 e 24, desenhava pontos, se estivesse entre 25 e 50, desenhava cruces, e assim sucessivamente. No entanto, as transições entre intervalos nesta abordagem eram extremamente repentinas, por isso, tentou-se encontrar uma forma de colocar as formas a alterarem gradualmente consoante a amplitude. Assim, desenvolveu-se um algoritmo para desenhar cada forma de acordo com uma probabilidade que varia ao longo da amplitude. Este algoritmo permite que exista maior probabilidade de o programa desenhar pontos em amplitudes mais baixas, e, à medida que a amplitude aumenta, essa probabilidade decresça, aumentando a probabilidade de desenhar cruces, depois círculos e, nas amplitudes mais elevadas, rectângulos. O algoritmo divide-se em duas partes: **i**) cálculo do **factor de crescimento** de cada forma ao longo da amplitude (é necessário calcular este factor primeiro, pois a probabilidade de desenhar cada forma é relativa à probabilidade das outras formas e não absoluta) e **ii**) cálculo da **probabilidade** de cada forma ser desenhada ao longo da amplitude. As funções encontradas para o factor de crescimento (com valores definidos entre 0 e 100) de cada forma foram as seguintes, sendo **A** a amplitude actual:

Ponto	$= 100 * \sin(\pi/100 * A + \pi/2)$
Cruz	$= 100 * \sin(\pi/75 * A)$
Círculo	$= 100 * \sin(\pi/75 * A - \pi/3)$
Rectângulo	$= 100 * \sin(\pi/100 * A - \pi/2)$

O gráfico relativo a estas fórmulas é representado no Gráfico 2. No eixo dos YY, pode ver-se o factor de crescimento de cada uma das formas (entre 0 e 100) e no eixo dos XX, a amplitude actual.

A fim de obter a probabilidade de cada forma em cada momento, é necessário mapear os valores do factor de crescimento das quatro formas entre 0% e 100%. Para tal, efectua-se a seguinte regra de três simples para cada uma das formas, a fim de obter a probabilidade de cada uma, sendo

V o valor resultante da função do factor de crescimento de cada forma na amplitude actual e S a soma dos quatros valores na amplitude actual:

$$\text{Probabilidade} = (V*100) / S$$

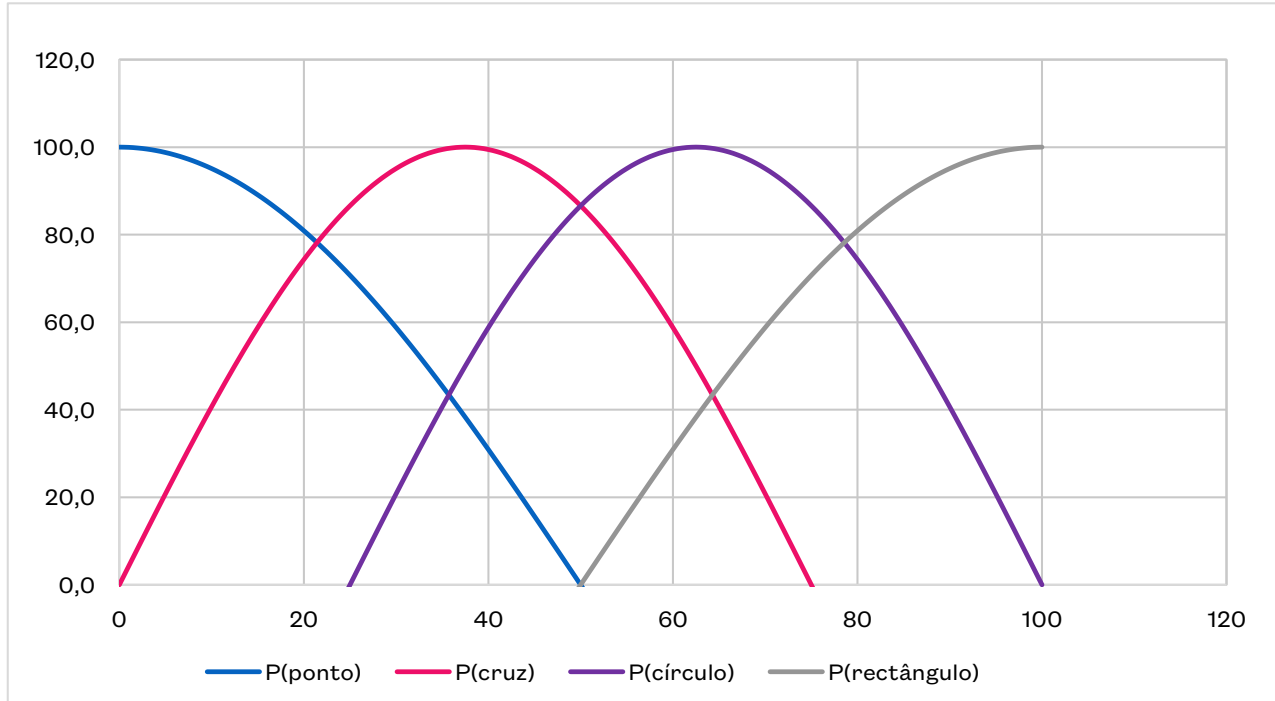


Gráfico 2 — Factor de crescimento de cada forma em função da amplitude.

Desta forma, assumindo que o factor de crescimento das diferentes formas na amplitude de 20 dB são os que constam na Tabela 2, aplicando a regra de três simples a cada uma das formas, obtêm-se os correspondentes valores de probabilidades.

Forma	Factor de crescimento	Probabilidade (%)
Ponto	81	52.3
Cruz	74	47.7
Círculo	0	0
Rectângulo	0	0

Tabela 2 — Probabilidade de cada forma em função do seu factor de crescimento para 20 dB.

Com esta funcionalidade, é possível, em cada valor de amplitude, ter uma, duas ou três diferentes ao mesmo tempo, tornando o aspecto das ilustrações mais diversificado.

O objecto `[env~]` do *Pure Data* tem tendência a devolver valores de amplitude maioritariamente entre 40 dB e 90 dB, apesar de, na teoria, este valor ir de 0 dB a 100 dB. A fim de garantir um melhor funcionamento da visualização deste atributo, efectuou-se um mapeamento dos valores devolvidos pelo objecto `[env~]` para o intervalo entre 0 e 100: sempre que o valor devolvido for menor ou igual a 40 dB, a amplitude assume o valor de 0 dB, sempre que for maior ou igual a 90 dB, a amplitude assume o valor de 100 dB e para cada valor entre 40 dB e 90 dB calcula-se o valor correspondente no intervalo entre 0 dB e 100 dB.

As figuras seguintes apresentam os resultados visuais correspondentes a alguns valores de amplitude devolvidos pelo objecto `[env~]` do *Pure Data*.

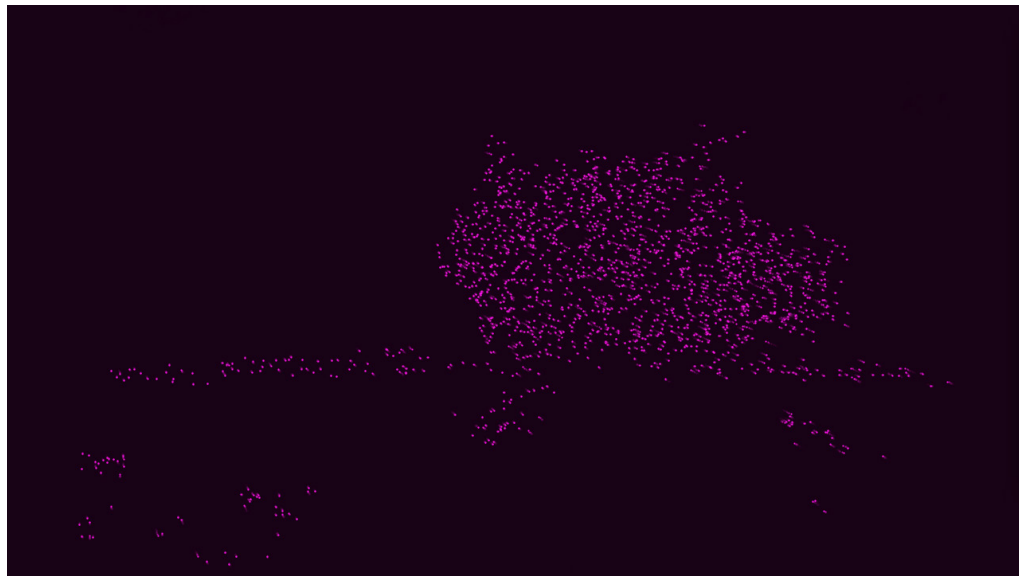


Figura 66 — *Frame* de quando a amplitude é aproximadamente 40 dB no *Pure Data* e 0 dB após o mapeamento.

Quando a amplitude é de aproximadamente 40 dB no *Pure Data* e 0 dB após o mapeamento (Fig. 66), pode verificar-se que existem apenas pontos na figura, pois a probabilidade de existir pontos no valor mínimo de amplitude é de 100%, enquanto a de existir qualquer uma das outras formas é de 0%.

Quando a amplitude é de aproximadamente 68 dB no *Pure Data* e 56 dB após o mapeamento (Fig. 67), já existem mais cruzeiros do que pontos, não existindo ainda outras formas.

Quando a amplitude é de aproximadamente 75 dB no *Pure Data* e 70 dB após o mapeamento (Fig. 68), existem mais círculos, ainda com pontos e cruzeiros em simultâneo, mas em menor quantidade.

A amplitude de aproximadamente 86 dB no *Pure Data* e 92 dB após o mapeamento (Fig. 69) já é muito próxima do máximo da amplitude, pelo que existem quadrados em maioria, alguns círculos e muito poucos pontos e cruzeiros.

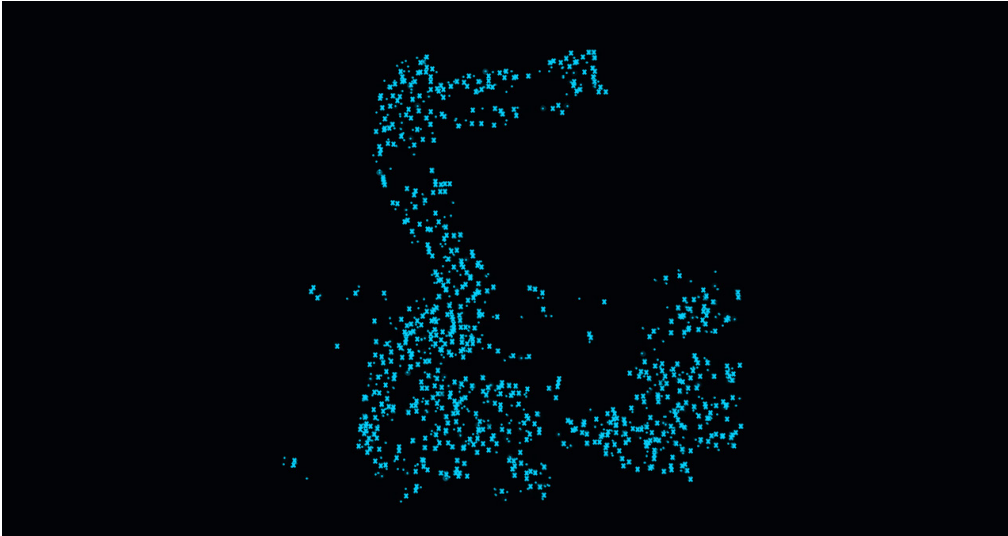


Figura 67 — *Frame* de quando a amplitude é aproximadamente 68 dB no *Pure Data* e 56 dB após o mapeamento.

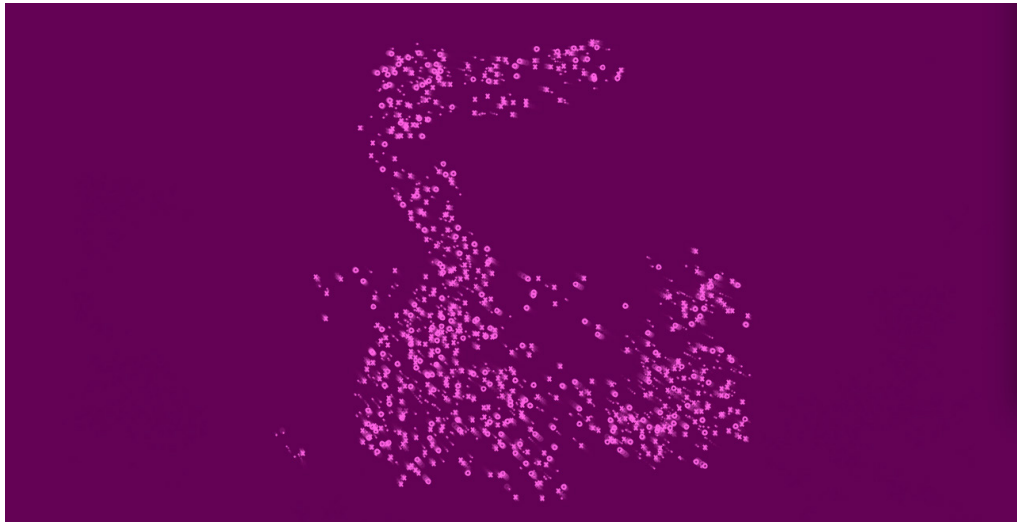


Figura 68 — *Frame* de quando a amplitude é aproximadamente 75 dB no *Pure Data* e 70 dB após o mapeamento.

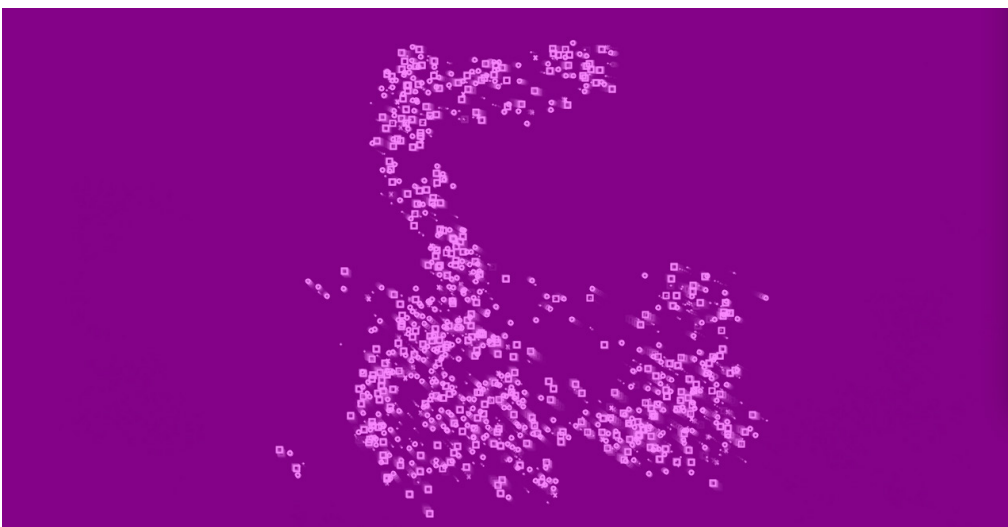


Figura 69 — *Frame* de quando a amplitude é aproximadamente 86 dB no *Pure Data* e 92 dB após o mapeamento.

Spectral Centroid

O *spectral centroid* corresponde ao centro de gravidade do espectro da magnitude, resultante de uma análise *STFT* (*short-time Fourier transform*) (Yang & Chen, 2011, p. 44). Quanto maior é o seu valor, mais “brilhante” é a textura do áudio. Como pode constatar-se nos resultados apresentados a seguir, as canções com mais *Arousal* são também as que têm a média do *spectral centroid* maior:

- a) 2968 Hz
- b) 2500 Hz
- c) 1150 Hz
- d) 1800 Hz

Estes valores foram obtidos a partir do objecto [*specCentroid~*] da biblioteca *timbreID* do *Pure Data* e correspondem à frequência média (em Hz) onde se situa o *spectral centroid*. Existe uma boa distinção entre os valores das quatro canções: a canção a) é a que possui maior quantidade de frequências altas e a c) é a que possui menor quantidade de frequências altas. Assim, estes valores também são concordantes com as conclusões retiradas dos diversos estudos analisados. O *Arousal* é bastante perceptível na análise deste atributo, portanto, o seu valor pode ser mapeado com uma variável visual que represente o *Arousal* da música. Ao observar os valores médios de *spectral centroid*, percebeu-se que os valores se encontravam sobretudo entre 500 Hz e 3100 Hz, por isso, estes foram os valores mínimo e máximo de mapeamento com a variável visual seleccionada. Escolheu-se, então, o *spectral centroid* para representar o *color hue* do fundo ou dos pontos, dependendo do brilho do fundo em cada momento: considerando o brilho do fundo a crescer desde 0 até 100, a saturação dos pontos decresce de 100 para 0, ou seja, quando o fundo tem o brilho no máximo, os pontos ficam brancos e o fundo com a sua *color hue*, quando tem o brilho no mínimo, os pontos ficam com a sua *color hue* e o fundo fica preto. Desta forma, existe sempre um contraste de cor entre o fundo e os pontos. Por intuição, associou-se músicas calmas à cor azul escura — pois é uma cor relacionada com tristeza, mas também com calma e paz, podendo a música ser depressiva ou pacífica/contente e não destoar — e as músicas energéticas à cor vermelha/laranja — pois é uma cor relacionada com raiva, mas também com excitação e exuberância, podendo a música ser raivosa/ansiosa ou alegre/animada e não destoar. Assim, na escala de 360° que representa os diversos valores de *color hue* (Fig. 70), optou-se por se iniciar o mapeamento nos 200°, que corresponde ao início do azul escuro e terminar nos 60°, após voltar ao início, que corresponde ao fim dos tons quentes. Desta forma, as cores verde e cyan foram postas de lado mais uma vez por um motivo subjectivo, pois não têm uma relação intuitiva com as emoções escolhidas — tristeza/calma/paz e raiva/excitação/exuberância.

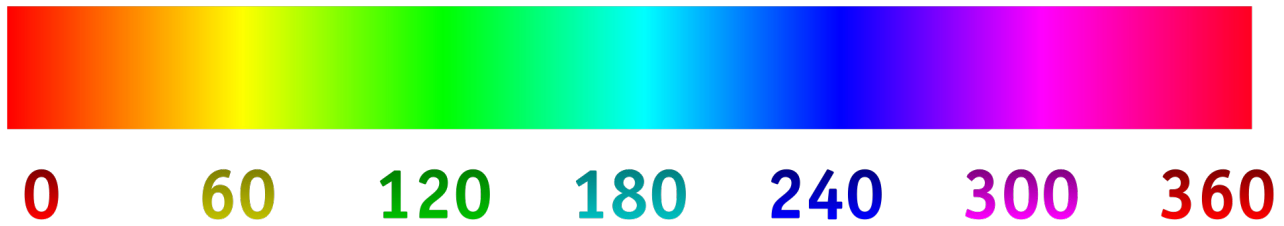


Figura 70 — Escala de 360° de *color hue*.

Numa primeira tentativa, mapeou-se a média do *spectral centroid* desde o início da canção até ao instante actual à cor do fundo. No entanto, quanto mais valores existem no cálculo da média, mais o valor da média estabiliza, pelo que, ao início a cor variava muito, mas num instante mais avançado da canção, mesmo que a intensidade da música variasse muito, tal não era tão perceptível na média. Para corrigir isto, optou-se por utilizar a média apenas dos últimos 150 *frames* (cerca de 8 ou 9 segundos) em cada instante, eliminando os valores anteriores a esses 150 *frames*. Desta forma, a *color hue* representa melhor a intensidade da música em cada trecho, podendo uma canção calma ter o fundo azul escuro, mas também vermelho/laranja em partes mais intensas da música. Para ilustrar este raciocínio, foi realizado um esquema apresentado na Figura 71, considerando a música *Take Me From Here*, já referida:

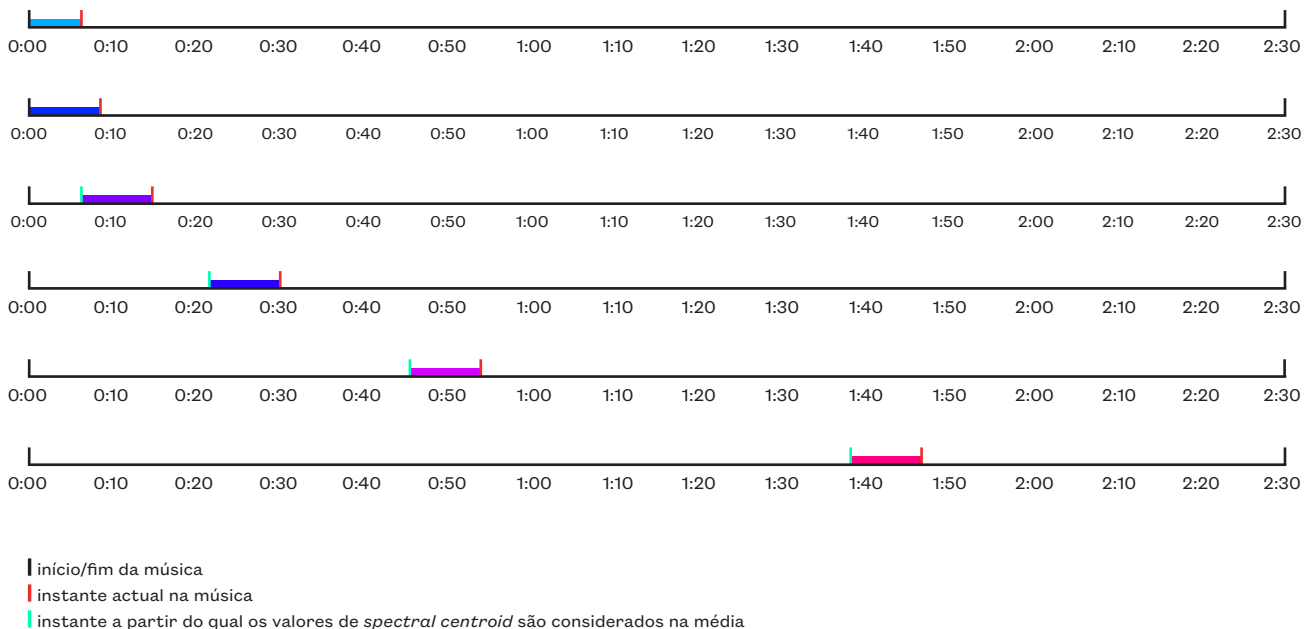


Figura 71 — Esquema da *color hue* considerando o intervalo de tempo definido para o cálculo da média do *spectral centroid*.

O mesmo raciocínio foi aplicado ao desenho das formas dos pontos, já descrita na secção anterior, mas com uma janela de frames consideravelmente menor (25 *frames*), pois, ao utilizar o valor da amplitude actual

directamente, as variações das formas surgiam muito repentinamente. A utilização da média dos últimos 25 *frames* neste cálculo permitiu que a variação das formas se tornasse mais suave.

Spectral Flux

O *spectral flux* corresponde à quantidade de mudança nas magnitudes do espectro sonoro entre *frames* sucessivos (Tzanetakis & Cook, 2002). Ao calcular a média destes valores, obtidos pelo objecto [specFlux-] da biblioteca *timbreID* do *Pure Data*, é possível observar claramente quais são as canções com mais diferenças espectrais:

- a) 215,4
- b) 399,2
- c) 70,1
- d) 99,2

Mais uma vez, as canções que possuem maior *Arousal* são as canções com valores mais elevados de *spectral flux*. Existiam diversas alternativas de variáveis visuais para serem influenciadas por este atributo. Pensou-se, por exemplo, em utilizá-lo para definir a quantidade de mudança (positiva ou negativa) que uma variável visual sofria, como, por exemplo, a transparência de algum elemento gráfico presente no programa. No entanto, este atributo acabou por ser visualmente metaforizado, ao colocar-se o tamanho do arrasto dos pontos tanto maior quanto a diferença espectral em cada *frame*. Se o valor actual de *spectral flux* for maior do que o seu valor médio ao longo da música, o grau de transparência do fundo incrementa, se for menor do que a sua média, o grau de transparência do fundo decrementa. Assim, quando a canção está mais intensa — ou seja, com maiores mudanças no espectro sonoro entre *frames* — os pontos têm mais arrasto, quando está mais calma, os pontos têm menos arrasto (mantendo sempre, no entanto, algum arrasto mínimo).

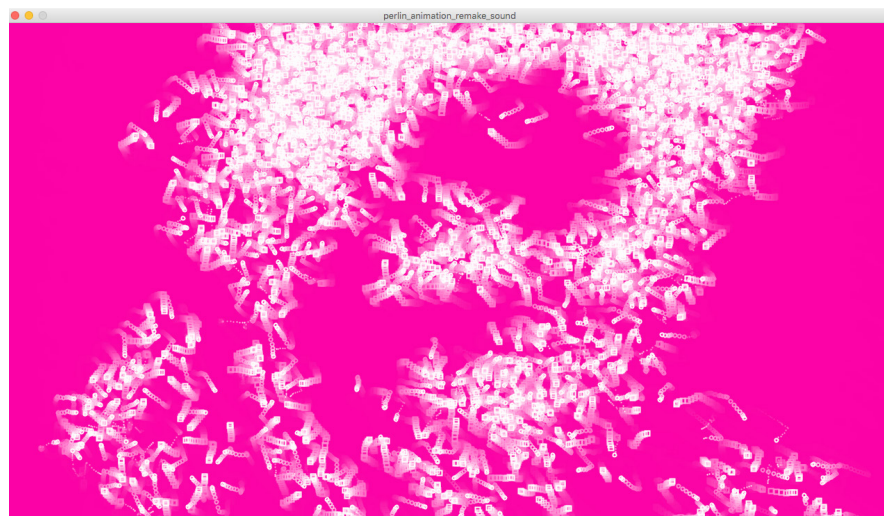


Figura 72 — *Frame* em que é possível observar o arrasto das partículas.

No final dos cinco módulos descritos até aqui, obtém-se um *output* animado como o que se apresenta no seguinte *link*:

<https://youtu.be/1Q-X1bijDg>

5.9 Atributos sonoros estudados mas não utilizados

Os três atributos sonoros descritos na secção anterior foram seleccionados após uma pesquisa detalhada de diversos atributos sonoros descritos nos vários estudos analisados para este trabalho. Nesta secção apresenta-se os atributos que não foram utilizados e os motivos pelos quais foram excluídos. Também para esta secção foram utilizadas as quatro músicas **a)**, **b)**, **c)** e **d)** para a análise de cada atributo.

Pitch

O *pitch* corresponde à frequência ou altura da nota do sinal sonoro.

Já desde há muito tempo que se tenta criar associações entre luz e som, nomeadamente entre *color hue* ou *brightness* e *pitch* ou *loudness*. Por exemplo, **i)** Aristóteles escreveu, em 440 a.C., sobre o paralelismo que existe entre a harmonia dos sons e a harmonia das cores compostas, **ii)** em 1704, Newton tentou expressar uma relação explícita entre sete cores primárias e sete notas da escala musical e em 1725, **iii)** Castel criou um *color organ*, em que notas musicais particulares tinham uma cor particular correspondente: dó com azul, ré com verde, sol com vermelho, etc. (Marks, 1974). De facto, é tentador associar as percepções visuais com as percepções audíveis, pois, muitas vezes, têm atributos semelhantes, como é o caso da “qualidade” e “intensidade”: a manifestação visual da “qualidade” é a *color hue*, enquanto a manifestação audível é o *pitch*; a “intensidade” é geralmente associada ao *brightness* na visão e à *loudness* na audição (*Idem*). No entanto, não existe uma direcção única nestas relações, pois também existem estudos que estabelecem uma relação entre *brightness* e *pitch*, por exemplo. Marks (1974) conclui o seguinte no capítulo da discussão:

When subjects are asked to match pure tones to gray surfaces, they map increasing pitch (sound frequency) onto increasing surface brightness (luminance) so long as loudness (sound pressure) remains constant. When pitch remains constant, most subjects map increasing loudness onto increasing brightness, but some map loudness into darkness.

Num estudo mais recente de Grill e Flexer (2012) foi proposta uma estratégia de visualização para representar atributos relevantes de sons. O mapeamento realizado para o *pitch* do som revela uma opção interessante em que se combina o *brightness* com a *color hue*, em que a escala de *high* para *low* começa em *bright yellow* e acaba em *dark red*.

Ao realizar-se experiências de representação visual do *pitch* a partir da alteração do *color hue*, utilizando a canção seleccionada, concluiu-se que uma representação por secções da música (conjuntos de *frames* consecutivos) poderia conceder mais firmeza à representação visual deste atributo sonoro e uma maior percepção da sua evolução ao longo da música, ao invés de uma representação para cada *frame* da música. Assim, optou-se por analisar a média deste atributo ao longo da música. O objecto [fiddle-] do *Pure Data* devolve o valor da nota MIDI estimada do sinal sonoro em cada *frame*. Ao calcular a média das notas MIDI de cada canção (numa escala de 0 a 127) e respectiva frequência, obtiveram-se os seguintes resultados:

- a) 61,25 — 281,21 Hz
- b) 61,66 — 287,95 Hz
- c) 56,80 — 217,47 Hz
- d) 59,67 — 256,69 Hz

Analisando estes resultados, pode concluir-se que a média das frequências é mais elevada nas canções com mais *Arousal*, **a)** e **b)**. No entanto, os valores não são suficientemente contrastantes para concluir com certeza que quanto maior o *pitch* médio da música, maior o seu *Arousal*. Os valores obtidos na análise, por exemplo, do *spectral centroid* — que corresponde ao centro de gravidade do espectro da magnitude, o que significa que quanto mais frequências altas existirem no som analisado, mais elevado é este valor — ou do *spectral rolloff* — que corresponde à frequência que representa o ponto do espectro sonoro abaixo do qual está concentrada uma percentagem (normalmente 85%) da distribuição da magnitude (energia), o que significa, mais uma vez, que quanto mais frequências elevadas existirem no som analisado, mais alto se encontrará este limite — são consideravelmente mais contrastantes entre músicas com mais e menos *Arousal*. Assim sendo, optou-se por utilizar o *spectral centroid* em vez do *pitch* para realizar um mapeamento com a *color hue*, não só por ser mais claro a nível computacional, mas também por manter visualmente a percepção do ouvido humano na variação da média do *pitch* ao longo da música. Tendo em conta que os estudos analisados sobre a relação que existe entre a cor e o *pitch* revelaram-se muito inconclusivos, optou-se por utilizar a abordagem subjectiva já descrita para a associação da altitude da frequência do *spectral centroid* com o nível de *color hue*, tendo por base a temperatura da cor.

Ao combinar as associações realizadas entre o *spectral centroid* e a *color hue* e entre a amplitude e o *brightness*, como sugerido em alguns dos estudos analisados, conseguiu-se obter um resultado bastante satisfatório na representação visual do *Arousal* da música e da variação da média do *pitch* e da intensidade (amplitude) ao longo da música.

Tempo

O tempo corresponde aos batimentos por minuto sobre os quais uma música se desenvolve. Quanto maior for este valor, mais rápida é a música.

A detecção computacional do tempo de uma música ainda é bastante problemática devido a diversos factores, entre eles:

- i) Em canções sem uma batida audivelmente definida, como, por exemplo, canções de piano e voz, é difícil encontrar um ponto de referência que permita calcular o tempo da música.
- ii) Não se pode guiar pelo ritmo da música para definir o tempo da música, porque um ritmo composto por batidas rápidas pode estar na mesma inserido num tempo lento. Além disso, muitas vezes existem contratempos no ritmo e a batida do tempo insere-se num instante de silêncio presente no ritmo.
- iii) Assumindo que se consegue calcular o tempo da música, é difícil para o programa definir, ao analisar uma música de, por exemplo, 120 bpm, se o tempo é de 60 bpm, 120 bpm ou 240 bpm, pois todos encaixam na música, mas nem todos correspondem ao andamento correcto da música.

Estes problemas foram também salientados pelo Prof. Dr. Rui Pedro Paiva na já mencionada reunião via *Skype*, e foram confirmados ao testar alguns mecanismos de detecção de tempo desenvolvidos para o *Pure Data*, nomeadamente o objecto [aubiotempo~] da biblioteca *pd-aubio*, o [tempo~] da biblioteca *timbreID* e outras *patches* desenvolvidas e partilhadas por utilizadores do *Pure Data*.

O objecto com resultados mais satisfatórios foi o [tempo~] do *timbreID* e foi utilizado para analisar o tempo das quatro canções. Como este objecto calcula constantemente o tempo da música, o valor varia muitas vezes. Assim, a fim de obter o valor mais provável para o tempo, aglomerou-se todos os valores devolvidos pelo objecto numa lista e calculou-se a moda. O valor com mais ocorrências corresponde, assim, ao valor mais provável. Como é pouco comum uma música ter um tempo inferior a 65 bpm, definiu-se uma regra no programa para que, se o objecto devolver um tempo inferior a 65 bpm, esse valor é multiplicado por dois. Esta regra permitiu aumentar a probabilidade de obter o múltiplo certo da música, no entanto, fragilizou um pouco mais a detecção do tempo, pois, apesar de raras, as canções com tempo inferior a 65 bpm existem e, após a aplicação desta regra, o tempo devolvido é muito mais rápido do que deveria ser. Estes foram os resultados obtidos:

- a) 118 bpm
- b) 112 bpm
- c) 88 bpm
- d) 112 bpm

Na teoria, canções com tempo mais rápido têm mais *Arousal*, no entanto, analisando estes resultados, as canções **b)** e **d)** têm exactamente o mesmo valor tempo, apesar de a **b)** ser muito mais energética do que a **d)**. Além disso, o objecto, apesar de ter conseguido detectar o tempo correcto da canção **c)**, demorou muito tempo da música a fazê-lo, pois esta canção não contém percussão e o ritmo é pouco perceptível.

Tendo em conta todas estas limitações, optou-se por não se utilizar este atributo na visualização da música. Ao mapear este atributo com alguma variável visual, poderia gerar-se ruído visual enquanto o objecto não devolvesse um valor fixo, ou poderiam gerar-se efeitos visuais completamente desadequados à canção em questão ao devolver valores errados.

Onset Frequency (Event Density)

O *onset* corresponde ao início de um evento (nota) na música. É possível detectar *onsets* através de vários objectos do *Pure Data*, nomeadamente o **[bonk~]**, o **[bark~]** da biblioteca *timbreID*, o **[aubioonset~]** da biblioteca *pd-aubio*, o **[essentiaRT~]** da biblioteca *essentiaRT*- (uma derivação da biblioteca *Essentia*, desenvolvida para análise de som em *C++*, mas para *Pure Data* e *Max/MSP*). Normalmente a *onset frequency* (ou *event density*) calcula-se pelo número de eventos (notas) num segundo.

As médias da *onset frequency* em cada uma das quatro músicas são apresentadas abaixo — o primeiro valor corresponde à média por segundo e o segundo valor à média por minuto:

- a) 6,781 — 406,86
- b) 5,868 — 352,08
- c) 6,452 — 387,12
- d) 6,031 — 361,86

Era de esperar que uma canção mais rápida e energética tivesse mais eventos ou notas num segundo do que uma canção mais lenta e calma. No entanto, nem sempre tal acontece, como se pode ver pelos resultados. A música **b)**, que é a segunda música com mais *Arousal*, é a que possui a menor média de notas por segundo/minuto e a canção **c)**, a mais calma entre as quatro, é a segunda canção com mais eventos ou notas por segundo/minuto.

Apesar de este atributo não servir de ajuda na representação visual do *Arousal* ou da *Valence* de uma música, poderia ser utilizado em *real-time* (em vez do valor da média ao longo da música), para determinar a ocorrência de algum evento visual sempre que detectasse uma nova nota. Apesar de ter sido desconsiderado nesta fase do trabalho, devido aos testes efectuados não terem tido resultados satisfatórios, pretende-se, no futuro, explorar com mais detalhe a forma como este atributo poderá influenciar a visualização final de uma música neste projecto.

Zero-crossing rate

A *zero-crossing rate* corresponde ao número de vezes que uma onda sonora atravessa o eixo no 0. Quanto maior for este valor, maior o nível de *noisiness* da onda sonora, o que significa que maior é o *Arousal*. Também para este atributo existem diversos objectos do *Pure Data*, nomeadamente, o [aubiozcr~] da biblioteca *pd-aubio* e o [zeroCrossing~] da biblioteca *timbreID*. Neste caso, utilizou-se este último e, ao calcular a média deste valor durante cada música, obtiveram-se resultados bastante contrastantes entre as canções a) e b) e as canções c) e d), demonstrando, assim, que esta medida funciona de forma correcta na obtenção do *Arousal* de uma música:

- a) 61,90
- b) 54,10
- c) 18,70
- d) 22,75

Um dos problemas deste atributo é o facto de não existirem limites de referência, nem inferior nem superior, para um eventual mapeamento com uma variável visual. Pretende-se, no futuro, explorar um mapeamento com uma variável visual que se relacione com o *noisiness* da música, como, por exemplo, a distorção da ilustração ou das formas individuais que a compõem. Uma possibilidade para obter limites aproximados para referência num mapeamento como este seria, tal como se fez para a detecção dos limites do *spectral centroid*, seleccionar canções que representem bem os dois extremos — canções com um elevado nível de *noisiness*, no caso do limite superior, e canções com o mínimo de *noisiness*, no caso do limite inferior —, analisar os valores médios mínimos/máximos ou os valores absolutos mínimos/máximos de cada uma e utilizar esses valores como mínimo e máximo do mapeamento. No entanto, nesta fase, optou-se por excluir este atributo, não só devido a esta limitação, mas também porque a legibilidade da ilustração com pontos, que já é frágil, poderia ficar comprometida quando aplicada a distorção de acordo com este atributo sonoro.

Spectral-Rolloff

O *spectral rolloff* é a frequência que representa o ponto do espectro sonoro abaixo do qual está concentrada uma percentagem (normalmente 85%) da distribuição da magnitude (energia). Quanto mais elevado for esse valor, mais agudo é o som geral (Yang & Chen, 2011). Tal como no atributo tímbrico anterior, consegue obter-se o *Arousal* de cada canção sem grandes dificuldades. Ao calcular a média dos valores devolvidos pelo objecto [specRolloff~] da biblioteca *timbreID* do *Pure Data*, obteve-se os seguintes resultados:

- a) 5632 Hz
- b) 4750 Hz
- c) 1620 Hz
- d) 3380 Hz

Mais uma vez, como se pode observar, as canções **a)** e **b)** são as que contêm o valor de *spectral rolloff* mais elevado, o que significa que são as que têm um maior nível de frequências agudas e, por isso, mais *Arousal*. Como este atributo consiste, basicamente, num limite, acredita-se que poderia fazer sentido explorar o seu mapeamento, em trabalho futuro, com o limite máximo de alguma variável visual já utilizada no programa, como, por exemplo, o limite máximo do *color hue* ou o limite máximo da luminosidade. No entanto, tal como acontece com o *spectral centroid* e a *zero-crossing rate*, também este atributo não contém limite máximo de referência, pelo que se teria de analisar canções que contenham elevados níveis de frequências altas, a fim de obter um limite razoável para este valor. Tendo em conta que se trata de um valor um pouco instável e que se trata de um mapeamento que limita um pouco a evolução da variável visual seleccionada, acabou por se descartar esta hipótese.

Roughness

Roughness é um fenómeno complexo que quantifica a percepção subjectiva da rápida modulação da amplitude de um som de frequência baixa (entre os 15 Hz e os 300 Hz, aproximadamente)⁶⁶ que provoca algum desconforto. Por este motivo, associa-se este conceito à consonância e dissonância musical e tem sido reconhecido como uma técnica de expressão na composição e performance de música (Villegas & Cohen, 2008). Segundo Yang e Chen (2011), qualquer nota que não encaixe na harmonia principal é considerada dissonante e, por isso, é possível estimar a *roughness* computacionalmente medindo a dissonância entre todos os pares possíveis de picos do espectro. Por fim, os autores concluem que este atributo poderá estar associado à *Valence*, pois as canções **b)** e **d)** são as que apresentam menor valor de *roughness*, sendo a **b)** a que tem o menor valor (Yang & Chen, 2011). Tendo em conta esta afirmação, considerou-se que inserir este atributo na análise das músicas seria importante neste projecto, pois poderia oferecer alguma ideia relativamente à emoção (positiva ou negativa) contida numa música. Para obter esta medição, optou-se por utilizar o *Roughometer*, um objecto desenvolvido por Villegas e Cohen (2008) especificamente para o *Pure Data*. Este objecto foi desenvolvido de acordo com o modelo de Vassilakis, detalhadamente descrito no artigo já mencionado que contém alguma documentação relativamente ao desenvolvimento do *Roughometer*.

Utilizando este objecto, obteve-se os seguintes resultados para cada uma das canções analisadas:

- a) 20,02
- b) 13,62
- c) 8,66
- d) 11,01

⁶⁶ University of Salford Manchester. <https://www.salford.ac.uk/research/sirc/research-groups/acoustics/psychoacoustics/sound-quality-making-products-sound-better/accordion/sound-quality-testing/roughness-fluctuation-strength> (data de consulta: 23.03.2019)

Tendo em conta estes resultados, continuam a ser as canções com mais *Arousal* que têm os valores mais elevados de *roughness*, e não as canções de *Valence* negativa, como sugerido por Yang e Chen. Assim, optou-se por não se utilizar este atributo na análise, pois há outros atributos que medem o *Arousal* de forma mais eficaz.

5.10 Módulo #6 — Possíveis aplicações do sistema num álbum de música

A fim de demonstrar algumas possibilidades de aplicação do *output* obtido no sistema desenvolvido ao longo deste projecto num possível álbum de música, desenvolveu-se várias possibilidades de grafismo com diversas abordagens. Para tal, criou-se um módulo #6 que é de utilização opcional por parte do utilizador: um programa igual ao do módulo #5, mas que exporta *frames* em intervalos de 60 em 60 *frames* a partir do momento em que surge o primeiro verbo ou substantivo na música. Posteriormente, criou-se um programa que corta as imagens numa proporção de 1x1 (quadradas). Por fim, um terceiro e último programa utiliza as imagens quadradas obtidas e cria uma nova imagem com as imagens dispostas numa grelha. Este *output* revelou-se bastante interessante, pois permite visualizar o aspecto visual geral da música e, conseqüentemente, a evolução das emoções presentes ao longo da música. Deste programa poderão surgir resultados diversos dependendo do estilo da música que serve de *input* ao primeiro programa. Na Figura 73 pode observar-se duas imagens exemplo geradas por este programa, uma a partir da versão baseada em fotografias e outra a partir da versão baseada em ilustrações.

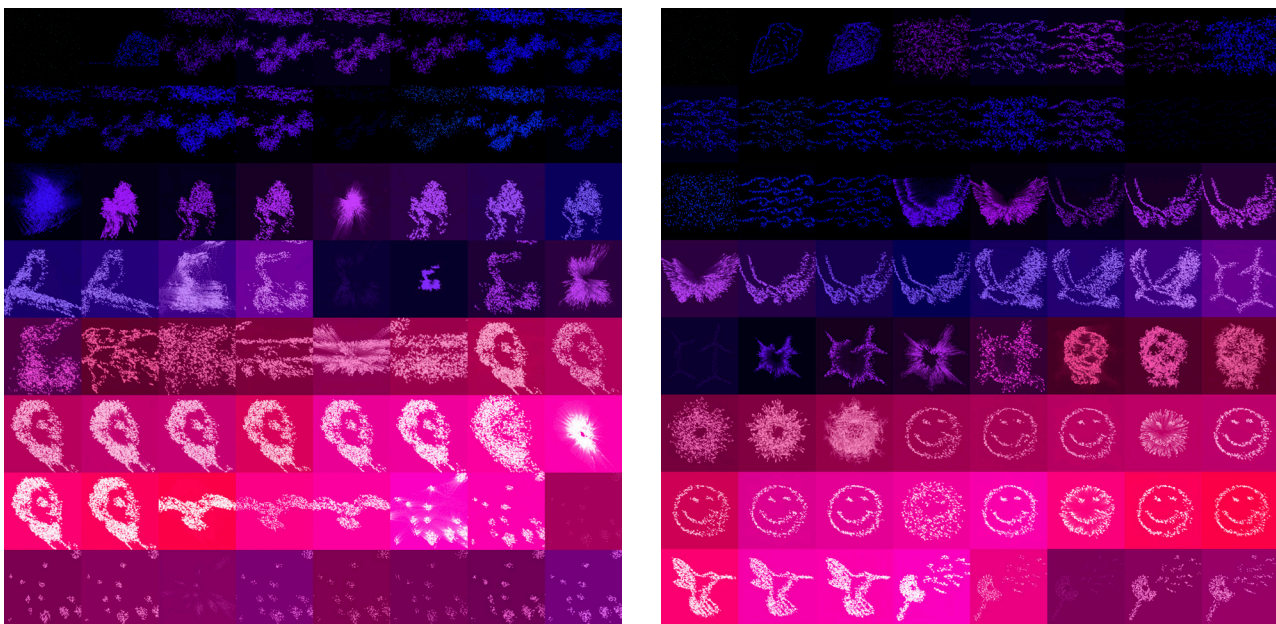


Figura 73 — À esquerda: versão do *output* baseada em fotografias. À direita: versão do *output* baseada em ilustrações.

Estes *outputs* poderão ser utilizados para criar capas para um álbum de música em diferentes materiais, como se demonstra na Figura 74.

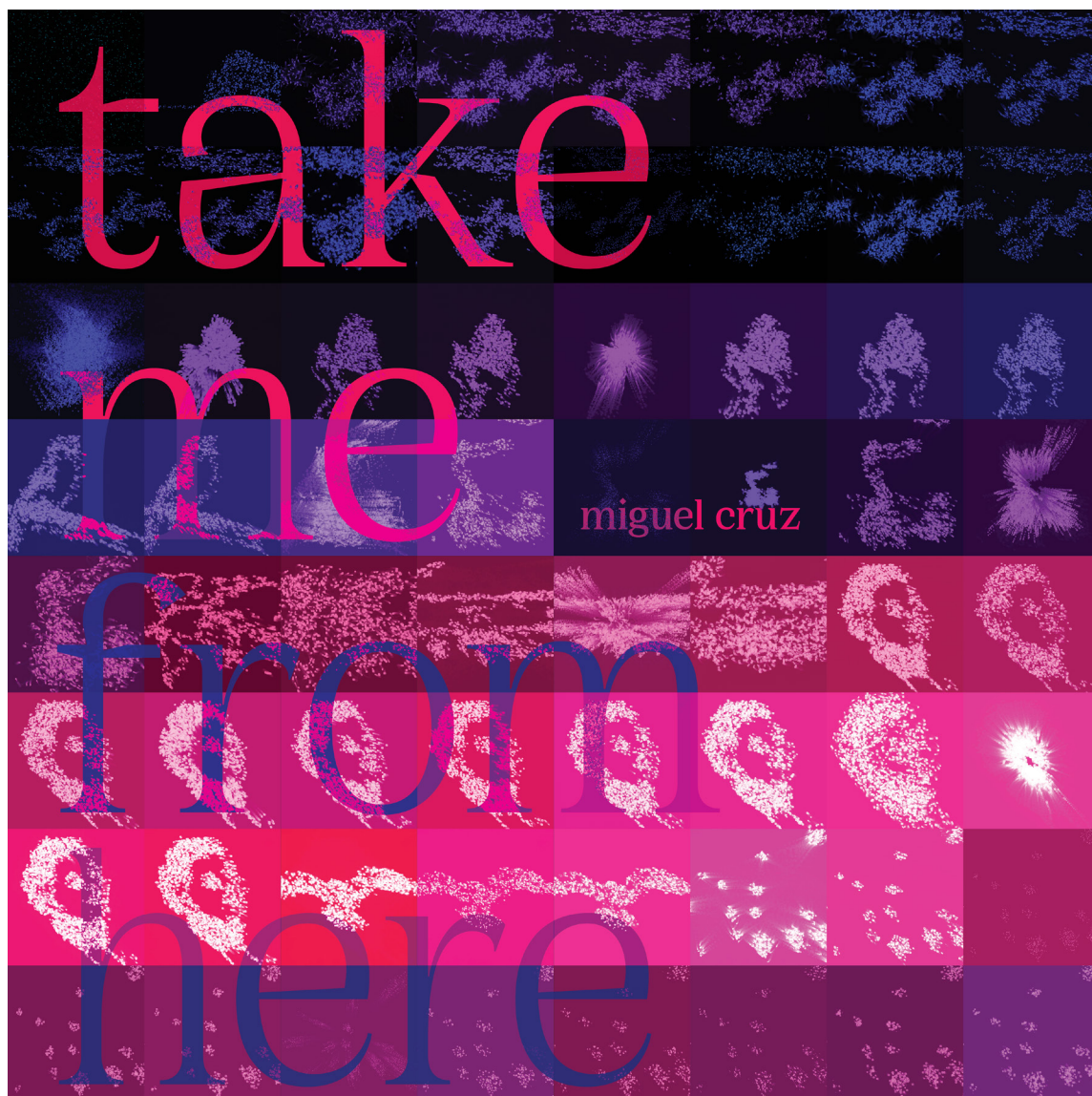


Figura 74 — Capa resultante da aplicação do *output* do módulo #6 e aplicações em vinil e cassette.

No entanto, as imagens geradas conferem bastante liberdade gráfica e podem ser utilizadas isoladamente para criar capas com diferentes *moods*. Na Figura 75 apresenta-se dois exemplos.

Como se pode observar, os resultados podem ser bastante diversificados para a mesma música, e o utilizador tem uma margem satisfatória para criar no seu registo próprio.



Figura 75 — Versões de capas utilizando *frames* isolados e que transmitem *moods* distintos.

Outra hipótese que esta abordagem permite é criar facilmente uma colecção de capas. As colecções de capas para um só álbum de música têm sido exploradas por diversos artistas, com o objectivo de tornar o disco num objecto de colecção. Um exemplo é o álbum *America* dos Thirty Seconds to Mars que foi destacado no capítulo dos **Casos de Estudo**. Apesar de todos os itens de uma colecção de capas serem constituídos pelo mesmo conteúdo, a diversidade no design confere um desejo de adquirir a colecção completa. Mesmo que isso não aconteça, atribui um leque mais elevado de possibilidades para os consumidores poderem escolher um item de acordo com os seus gostos pessoais. Experimentou-se, então, utilizar o sistema desenvolvido neste trabalho para criar uma colecção de capas para um hipotético *EP* com a canção *Take Me From Here*. Criaram-se dez capas com a mesma base tipográfica, mas com imagens diferentes (e, conseqüentemente, cores diferentes) criadas a partir do módulo #6 do sistema. A última contém as outras nove capas dispostas numa grelha, utilizando o programa desenvolvido no módulo #6 (Fig. 76). Esta colecção poderia ser aplicada em diversos materiais diferentes, como se demonstra na Figura 77.



Figura 76a — Coleção de capas criadas a partir do *output* gerado pelo sistema.

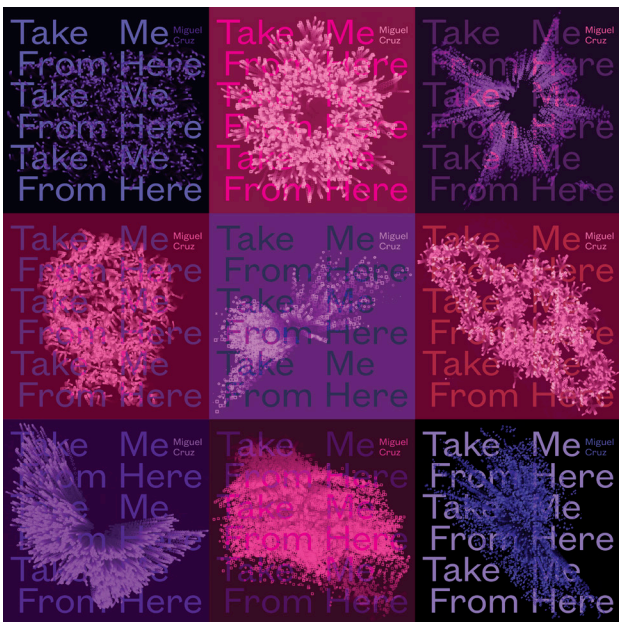


Figura 76b — Coleção de capas criadas a partir do output gerado pelo sistema.



Figura 77a — Possíveis aplicações das capas em diferentes materiais.



Figura 77b — Possíveis aplicações das capas em diferentes materiais.



Figura 77c — Possíveis aplicações das capas em diferentes materiais.

Para se poder visualizar o aspecto gráfico, sobretudo no que diz respeito às cores (associadas a características do som), bem como para se ter noção da evolução da dinâmica da música, seleccionou-se a *Feeling Good* em duas versões diferentes — uma original da Nina Simone e outra dos Muse. O resultado (Fig. 78) revela a diferença imprimida pelos autores em cada versão. De notar que nesta experiência não se teve em conta a letra da música, pois pretendia-se apenas comparar o aspecto geral da composição gráfica do *output*. É interessante verificar que existe uma diferença significativa entre os *outputs* das duas versões da mesma música. Além disso, é possível observar que as cores presentes nestas duas versões diferem das cores presentes na *Take Me From Here*.

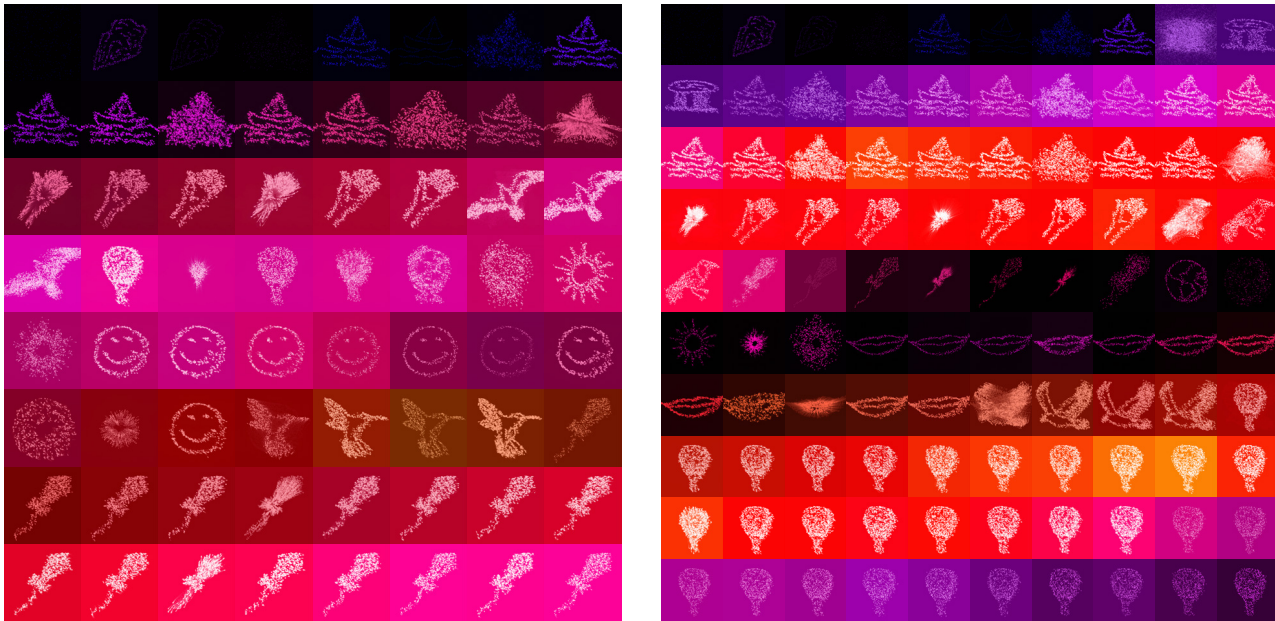


Figura 78 — À esquerda: *output* da *Feeling Good* na versão de Nina Simons. À direita: *output* da *Feeling Good* na versão dos Muse.

5.11 Módulo #7 — Aplicação do sistema nos efeitos visuais de um concerto

Neste módulo adaptou-se o programa desenvolvido no módulo #5 de modo a que reagisse ao som captado por um microfone em *real-time* em vez do sinal sonoro de um ficheiro de áudio de uma música. Tendo em conta que o sistema, de momento, está desenvolvido com instantes fixos de surgimento das ilustrações ao longo da música, o utilizador tem de tocar e cantar a música num determinado tempo para que, ao cantar, as ilustrações coincidam com as palavras presentes na letra da música.

Para efeitos de demonstração, montou-se um possível cenário de concerto, com algumas limitações, não só de espaço, mas também de material, pelo que a luz não é a ideal.

O vídeo de demonstração pode ser encontrado no seguinte *link*:

<https://youtu.be/gbcGyxsUGvo>

Como se pode observar, os efeitos visuais do concerto têm uma ligação visual com os elementos gráficos desenvolvidos como exemplo no módulo anterior. Desta forma, todos os materiais e eventos associados a um álbum de música, estáticos ou animados, estão inseridos num só sistema gráfico em que cada elemento funciona isoladamente e em conjunto com os restantes.

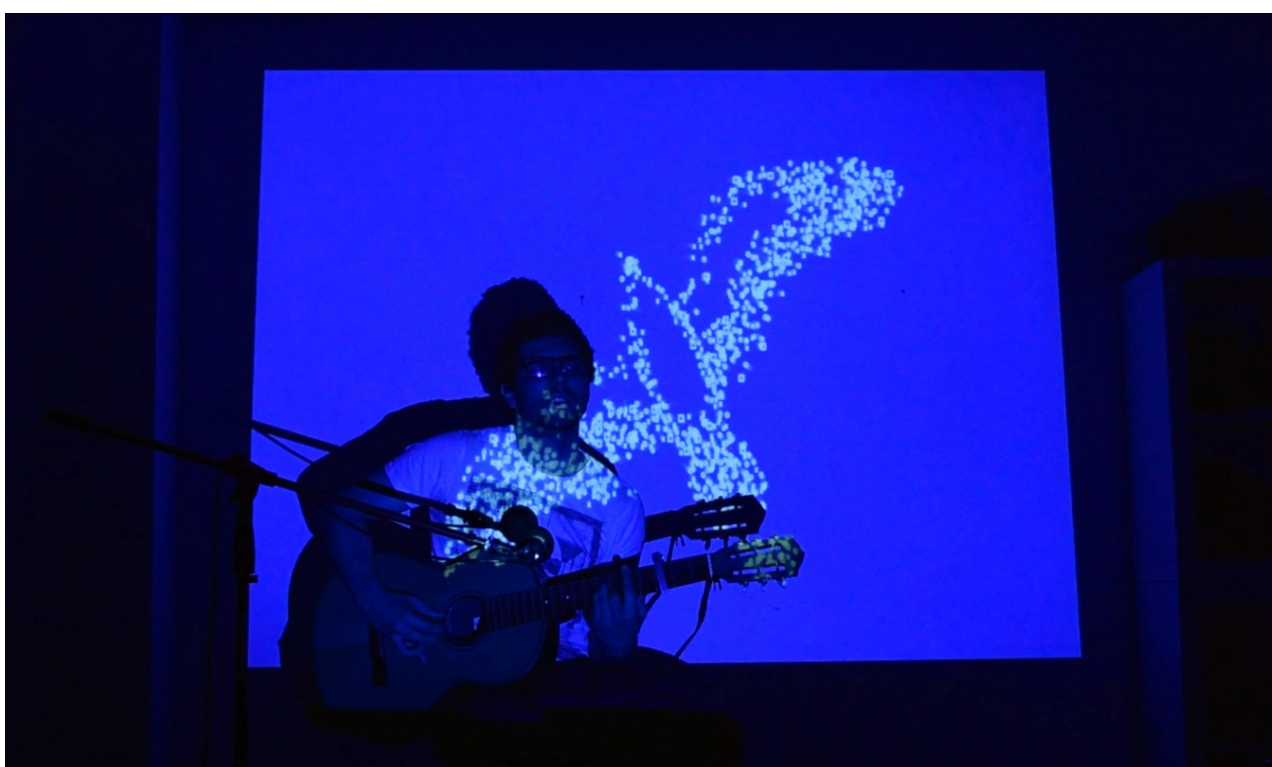
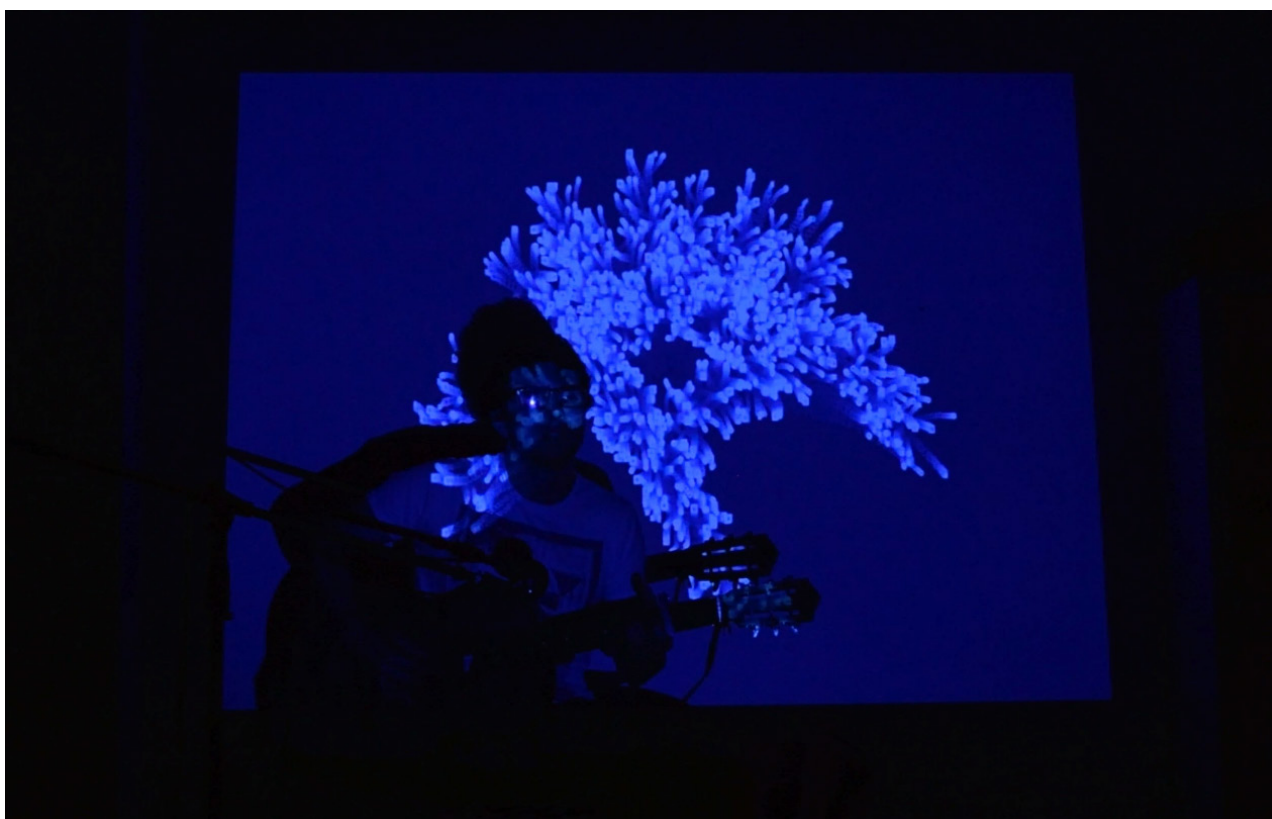


Figura 79 — Simulação de concerto para demonstração dos efeitos visuais gerados pelo sistema.

5.12 Discussão de resultados

O sistema resultante deste projecto partiu de um leque de problemáticas levantadas na primeira parte desta dissertação e acredita-se que todas elas foram exploradas durante o desenvolvimento do trabalho, obtendo-se soluções satisfatórias para cada uma. Desta forma, pensa-se que o produto final se revelou bastante versátil, no sentido em que pode ser aplicado em diversos contextos e de variadas formas.

No capítulo **Contextualização Histórica** mencionou-se uma crescente desintegração dos álbuns de música conceptuais, devido ao domínio dos *singles* e das *playlists* no consumo de música actual. Ao utilizar o sistema resultante deste trabalho, é possível unificar todas as canções de um álbum por meio do design e da multimédia, integrando todos estes elementos estáticos e animados num só vocabulário visual, fortalecendo, assim, o conceito de álbum conceptual, ou até mesmo, eventualmente, abrindo novas portas para a produção de álbuns conceptuais.

Além disso, todos os materiais e eventos associados ao álbum tornam-se reconhecíveis como pertencentes a um conjunto, funcionando também isoladamente, o que vai ao encontro da actual necessidade de pensar nos álbuns gravados e nos respectivos concertos de modo integrado.

O sistema parte sempre da letra, que contém a mensagem que o artista pretende transmitir e que constitui o motivo principal da visualização e, posteriormente, insere o som, que é o que molda a emoção da interpretação da mensagem, no controlo do aspecto gráfico geral da visualização. Desta forma, pensa-se que a representação visual da música criada pelo sistema enaltece os dois elementos que se considera serem os mais importantes numa música: a sua mensagem e a emoção nela contida. Esta capacidade permite combater a desvalorização das letras das músicas que tem acontecido tanto por parte dos consumidores como pelos próprios artistas, o que constitui outra problemática abordada na contextualização histórica desta dissertação.

Acredita-se que o sistema tenha alcançado o objectivo definido inicialmente de reunir capacidades presentes em três dos projectos mencionados no capítulo **Casos de Estudo** desta dissertação num só sistema: i) *Glitchometry*, que permite a obtenção de uma imagem estacionária modelada por som visualmente apelativa, ii) *Communion*, que apresenta a capacidade de reacção de uma imagem a um som produzido em *real-time* e iii) *Soundfield*, que possui a capacidade de reacção de imagem a um som captado em *real-time* a partir de interacção.

Em termos de aplicações possíveis, o mesmo sistema permite gerar *outputs* estáticos e animados que podem ser aplicados posteriormente em diferentes materiais e contextos, conferindo alguma liberdade ao utilizador durante a sua aplicação. É possível utilizar imagens estáticas geradas pelo sistema para criar uma imagem gráfica para um *single* ou um álbum, digital ou físico. Esta imagem gráfica pode adquirir alguma riqueza através de diferentes versões, podendo até ser composta por colecções de capas, como tem acontecido em alguns álbuns lançados recentemente, nomeadamente

o *America* dos Thirty Seconds to Mars. É possível utilizar o *output* animado do sistema — na versão em que o *input* é um ficheiro de áudio — como videoclipe de uma música ou — na versão em que o *input* é o microfone — como efeitos visuais da mesma música num concerto. Aplicado noutros contextos, é ainda possível utilizar o sistema como gerador de capas para os discos produzidos por uma editora discográfica, seguindo uma lógica de vocabulário visual semelhante à da *ECM Records*, da *4AD Records* ou da *Blue Note Records* e adicionando uma componente automática que pode facilitar o trabalho, um pouco como acontece na *Tonboutique Records*, apesar de numa abordagem distinta.

Por outro lado, o projecto desenvolvido apresenta ainda alguns aspectos que poderiam ser melhorados e que abrem portas para exploração futura de abordagens diferentes das que foram adoptadas.

No módulo #1 do sistema filtra-se a letra de modo a recolher apenas os substantivos e verbos, mapeando os substantivos às ilustrações e os verbos às animações das ilustrações. Esta sequência poderia ser estendida a outras POS (*parts of speech*), como, por exemplo, adjetivos, advérbios e interjeições, que poderiam representar algum aspecto visual do sistema, a fim de lhe conferir mais dinamismo. Ainda neste módulo, existe uma fragilidade no mapeamento realizado entre os verbos e as acções básicas seleccionadas, devido ao funcionamento da função `getDistance()` da biblioteca *RiTa* do *Processing*, que calcula a distância em número de nós percorridos na *WordNet* entre duas palavras, sendo necessária a existência de um nó parente comum, o que significa que não é possível calcular uma distância entre duas quaisquer palavras e, quando é possível, a precisão é consideravelmente baixa. O modelo *Word2Vec* apresenta uma possível solução para estes problemas e revela-se bastante aliciente para uma exploração futura.

Apesar do facto da criação ou recolha das imagens que servem de base para as ilustrações ser realizada pelo utilizador conferir um toque humano e pessoal ao sistema, existe uma interessante possibilidade de exploração futura no módulo #3 para o desenvolvimento de uma componente de recolha automática de imagens. Esta componente poderia basear-se, por exemplo, em API de plataformas de imagens como o *Unsplash* ou o *Flickr*.

Uma das principais fragilidades do sistema está associada à não existência de um reconhecimento automático das palavras cantadas, para que, posteriormente, seja possível atribuir a ilustração correcta no momento certo da música, podendo este ser variável de actuação para actuação. Esta limitação levou à existência de um módulo #2, que permite que o utilizador defina os instantes em que os substantivos e verbos presentes na letra surgem na música. Esta funcionalidade obriga a que o artista, quando está a actuar num concerto, tenha de tocar e cantar num tempo específico, sem grande margem para improvisação ou criação de versões diferentes. Poderia ser relevante explorar tecnologias de *speech recognition* para detectar as palavras cantadas e associá-las directamente à ilustração correcta no momento correcto. Uma das limitações ainda presentes nestas tecnologias, no entanto, é a dificuldade de reconhecimento de palavras num discurso cantado, que, além de não se manter sempre no mesmo tom, tem, muitas

vezes, prolongamentos de palavras para as encaixar na melodia ou ritmo da música.

Um problema existente no módulo #4 é o da frágil legibilidade existente em algumas ilustrações geradas. Por vezes, deve-se à escolha da imagem que serve de base, no entanto o facto de as ilustrações serem compostas apenas por pontos, embora facilite a modelação da ilustração de acordo com o som, retira um pouco da legibilidade, principalmente quando os pontos se movem. Também neste módulo se criou animações para os verbos presentes na letra. Tendo em conta que é impossível associar uma animação a cada verbo existente, optou-se por se mapear todos os verbos a quatorze acções básicas e uni-las em seis grupos, para os quais se associou uma animação. Criaram-se, assim, apenas seis animações diferentes. Quatorze animações, uma para cada acção básica, talvez conferisse um maior dinamismo e diversidade ao sistema.

O módulo #5 foi o que exigiu mais tempo de exploração e desenvolvimento. No entanto, apenas se utilizaram três dos nove atributos sonoros estudados neste trabalho, devido a algumas incoerências existentes entre os resultados obtidos no *Pure Data* e os estudos que foram analisados. Os resultados dos três atributos sonoros seleccionados coincidem com as conclusões retiradas nos estudos analisados sobre os mesmos atributos. Assim, espera-se aprofundar mais estes conhecimentos em trabalho futuro, a fim de adicionar mais atributos sonoros à componente de análise do som presente no sistema, conferindo um maior dinamismo e variedade no aspecto gráfico geral. Além disso, tendo em conta as conclusões presentes no estudo *Music Emotion Recognition* de Yang e Chen (2011), previu-se inicialmente que os nove atributos seleccionados estariam associados ou ao *Arousal* (se a música é calma ou energética) ou à *Valence* (se a música é negativa ou positiva) da música, segundo o modelo de Thayer (Thayer, 1989). No entanto, assumindo os valores obtidos em qualquer um dos nove atributos no *Pure Data*, concluiu-se que todos os atributos se relacionam apenas com o *Arousal*. Isto faz com que a *Valence* seja desconsiderada na representação visual da música, impedindo que seja possível distinguir visualmente uma canção energética e raivosa de uma canção energética e alegre, ou uma canção calma e triste de uma canção calma e pacífica. Desta forma, apenas a energia da música é representada visualmente, o que não invalida a representação visual das emoções contidas na música, mas torna-a menos rica a nível de diversidade. Um aspecto que é necessário ter em conta relativamente ao desenvolvimento deste módulo é a dimensão do *dataset*, ou seja, do conjunto de músicas utilizado para avaliar e analisar os atributos sonoros relativamente à sua relação com o modelo de Thayer. Foram utilizadas as quatro canções seleccionadas para demonstrar o estudo de Yang e Chen, cada uma relativa a um quadrante do modelo de Thayer. Isto permitiu avaliar os valores obtidos em cada uma das canções em termos de *Arousal* e de *Valence*. No entanto, tendo em conta a dimensão da amostra, pode afirmar-se que, se esta fosse maior, as conclusões obtidas poderiam usufruir de maior consolidação ou de maior significado.

Quanto à aplicação do sistema, existem ainda muitas possibilidades para o seu enriquecimento, como, por exemplo, o desenvolvimento de uma ferramenta que permita utilizar uma imagem gerada pelo sistema e inserir tipografia, a fim de gerar capas completas. Também existe a possibilidade de explorar a expansão do sistema para o contexto da *web*, permitindo, assim, a sua utilização no website do artista, um elemento bastante importante na promoção do álbum de música, e uma maior interacção com o público, como acontece no caso do *EP Prologue* de Emmitt Fenn. Tendo em conta que o sistema foi desenvolvido em *Processing*, poderá explorar-se a sua adaptação para *web* a partir da biblioteca *P5.js*, uma biblioteca que utiliza o mesmo vocabulário do *Processing* mas em *JavaScript*⁶⁷. Para a componente de análise de som, existe também uma biblioteca, *WebPd*, desenvolvida em *JavaScript*, que utiliza a *Web Audio API* e permite correr uma *patch* desenvolvida em *Pure Data* no *browser*⁶⁸.

67 P5.js. <https://p5js.org> (data de consulta: 22.06.2019)

68 GitHub. <https://github.com/sebpiq/WebPd> (data de consulta: 22.06.2019)

Conclusão

Capítulo VI

6.1 Trabalho futuro

Capítulo VI

Conclusão



Com esta dissertação pretendia-se enaltecer os aspectos que se consideram mais importantes numa música — a mensagem da letra e as emoções contidas no som — de modo a contribuir para uma melhor interpretação dessa música. Com este projecto foi possível criar um sistema que, utilizando a letra e características do som de uma música, gera *outputs* que criam um vocabulário visual, que contribui, não só para uma melhor interpretação da música, mas também para a unificação de todos os materiais e eventos associados a um álbum de música. Pensa-se que, deste modo, se recupera o conceito de álbum conceptual, trazendo uma nova abordagem que poderá ir ao encontro de outras tendências no consumo musical em que se pretenda associar a música à imagem.

Partindo de uma contextualização histórica da indústria fonográfica musical, identificou-se o problema para o qual se explorou soluções. De facto, desde o aparecimento da Internet, surgiram mudanças, de tal modo rápidas na indústria da música, que impediram um acompanhamento a igual velocidade nas estratégias de design e de marketing que permitissem a sua função de divulgação da música criada. Por outro lado, com as mais recentes formas de divulgação de música, nomeadamente com as plataformas de *streaming*, verifica-se que existe actualmente uma desintegração do conceito de álbum. No entanto, é de notar uma alteração de paradigma associado à indústria da música ao vivo em que os concertos, que antes serviam apenas para divulgar os álbuns lançados, têm sido, nos últimos tempos e nos casos de maior sucesso, parte integrante do conceito de álbum.

Tudo isto acontece num tempo em que os consumidores de música já se apropriaram da linguagem digital, sendo por isso as tecnologias digitais uma oportunidade para explorar um novo design que contribua para o marketing da música que se pretende fazer chegar ao público, de modo a que a mensagem passe com qualidade.

Justifica-se, assim, a pertinência do objectivo geral desta dissertação — criar um sistema que produza imagens que podem ser utilizadas para desenvolver uma imagem gráfica que realce a mensagem e emoções presentes num álbum de música. Ao mesmo tempo, pretende-se que o sistema integre visualmente todos os elementos associados ao álbum, nomeadamente, *packagings*, materiais digitais, materiais de promoção, entre outros, e ainda os efeitos visuais dos concertos, recuperando assim o conceito de álbum como mensagem e objecto artístico. Propõe-se uma estratégia para chamar a atenção do público, de modo a que sinta vontade de conhecer a música ao vivo ou em formato digital.

Com a formação na unidade curricular de Design de Som e com a pesquisa e exploração de ferramentas de design gráfico, design de interação e ainda de design generativo, foi possível obter conhecimentos teóricos e práticos relevantes para desenhar uma estratégia e encontrar uma solução ajustada ao problema. Para o aprofundamento do conceito e também da sua operacionalização, tiveram um papel de relevo os casos de estudo que se apresentam nesta dissertação e que foram seleccionados de acordo com a presença de algumas características que, de algum modo, estiveram presentes no projecto final desta dissertação. Algumas destas características são geração de imagem a partir de texto, geração de som a partir de texto, ilustração generativa, modelação de imagem de acordo com som. Incluem-se ainda exemplos práticos em que ferramentas de geração de capas foram úteis no mercado musical e alguns exemplos de álbuns que servem de inspiração e de referência para este trabalho. Em todos os casos apresentados, assinalam-se pontos fortes a ter em conta no desenvolvimento do projecto, mas também alguns pontos fracos que evidenciam aspectos que se tentou evitar ao longo do trabalho.

Depois de várias experiências com ferramentas digitais e tecnológicas, como, por exemplo, *Processing* e *Pure Data*, e também de explorações técnicas a partir das quais se identificaram algumas potencialidades, mas também constrangimentos, foi possível construir as estratégias para o desenvolvimento da componente prática do projecto e a exploração do conceito.

A construção do plano de trabalho, com a definição e sequência das tarefas a desenvolver, foi um recurso importante que contribuiu, tanto para o estudo teórico como para o processo criativo e técnico. Houve necessidade de reformulações ao longo do processo, em que se pretendia aprender o máximo, pois o principal objectivo era obter a melhor aprendizagem e o melhor produto possível.

O trabalho revelou-se como um constante desafio, desde logo pela ambição do problema que se pretendia explorar, mas também ao longo de todo o processo. A análise de texto presente na letra da música e o estudo das características do som susceptíveis de contribuir para uma interpretação das emoções contidas na música foram só por si de grande exigência. Durante a pesquisa sobre a teoria e a prática nestes domínios, houve a percepção de que esta é uma área ainda pouco explorada e onde não há grandes avanços feitos, sobretudo no que diz respeito a ferramentas de reconhecimento de discurso cantado. Também os estudos acerca das características do som passíveis de serem utilizadas na geração de ambientes visuais interpretativos da música não são muito consistentes. Foi, assim, necessário explorá-los a fundo, a fim de tomar decisões na selecção dos que respondessem adequadamente ao problema que se pretendia aqui abordar. O desafio da programação criativa foi muito interessante, tendo desencadeado a necessidade de aprofundamento de conhecimento das ferramentas utilizadas.

Com o sistema desenvolvido, pensa-se ter contribuído para i) facilitar a interpretação de um álbum de música com recurso a linguagem visual, ii) obter propostas interessantes, funcionais e visualmente apelativas de

imagens gráficas e efeitos visuais que poderão dar suporte à identidade de um álbum de música, a partir de um novo tipo de abordagem, **iii**) recuperar o conceito de álbum como portador de mensagem e também como objecto artístico.

De uma forma geral, o resultado final do projecto traz satisfação e confiança para explorar possíveis aplicações deste tipo de abordagem audiovisual num álbum de música. Acredita-se que as competências e conhecimentos adquiridos favorecem uma carreira como designer multimédia, colocando-a num nível que se revela bastante aliciante e desafiador.

Acredita-se ainda que, com esta dissertação, se contribuiu para o conhecimento e inovação na relação simbiótica entre os domínios da música, design e multimédia.

6.1 Trabalho futuro

O sistema criado neste projecto apresenta potencialidades que poderão, em trabalho futuro, ser ampliadas. Seria interessante explorar os seguintes aspectos:

- i) A partir do texto poder gerar imagens de modo eficaz, mas provenientes de uma base mais alargada de imagens.
- ii) Explorar o mapeamento com variáveis visuais (por exemplo, a distorção da ilustração ou das formas individuais) que se relacionem com outras características da música, como por exemplo, o *noisiness*.
- iii) O desenvolvimento de uma ferramenta que permita utilizar uma imagem gerada pelo sistema e inserir tipografia, a fim de gerar capas completas.
- iv) Utilizar os dados recolhidos a partir da música para determinar a tipografia e o *layout* a utilizar nas capas dos álbuns.
- v) Explorar a expansão do sistema para o contexto da *web*, permitindo **i**) a utilização do sistema no *website* do artista, para proporcionar uma maior interacção com o público e uma melhor promoção do álbum e **ii**) mudar a imagem associada a cada música do álbum em cada vez que é ouvida, (por exemplo, no telemóvel) associando o sistema às novas formas de ouvir música.

Referências

Bibliográficas

Capítulo VII

- 7.1 Bibliografia
- 7.2 Webgrafia

Capítulo VII

Referências Bibliográficas

7.1 Bibliografia

- Celma, Òscar e Lamere Paul (2011), “If You Like Radiohead, You Might Like This Article”, *AI Magazine*, vl. 32, no. 3, 57-66.
- Fu, Zhouyu, Guojun Lu, Kai Ming Ting e Dengsheng Zhang (2011), “A Survey of Audio-Based Music Classification and Annotation”. *IEEE Transactions on Multimedia*, vl. 13, no. 2, 303-319.
- Gamal, Ashraf El (2012), *The Evolution of the Music Industry in the Post-Internet Era*. Senior Thesis in Bachelor of Arts. Claremont McKenna College. Claremont, California.
- Grekow, Jacek e Rás, Zbigiew W. (1970), “Detecting Emotions in Classical Music from MIDI Files”, in J. Rauch, Z. W. Ras, P. Berka e T. Elomaa (eds.) (2009), *Foundations of Intelligence Systems*. Prague, Czech Republic: Springer, 261-270.
- Grill, Thomas e Flexer, Arthur (2012), “Visualization of Perceptual Qualities in Textural Sounds”, *Proceedings of the International Computer Music Conference (ICMC 2012)*, 589-596.
- Harrison, Cara (2014), *The visual album as a hybrid art-form: A case study of traditional, personal, and allusive narratives in Beyoncé*, Sweden: Lund University.
- Holt, Fabian (2010), “The economy of live music in the digital age”. *European Journal of Cultural Studies*, vl. 13, no. 2, 243-261.
- Jones, Steve e Sorger, Martin (1999), “Covering Music: A Brief History and Analysis of Album Cover Design”. *Journal of Popular Music Studies*, vl. 11, no. 1, 68-102.
- Lacerda, D. P., Aline Dresch, Adriano Proença e J. A. V. Antunes Júnior, (2013), “Design Science Research: método de pesquisa para a engenharia de produção”. *Gestão & produção*, vl. 20, no. 4, 741-761.
- Liu, Dan, Lie Lu e Hong-Jiang Zhang (2003), “Automatic Mood Detection from Acoustic Music Data”. China: Tsinghua University and Microsoft Research Asia.

- Long, Duri, Sanjana Gupta, Jessica B. Anderson, Brian Magerko (2017), “The Shape of Story: A Semiotic Artistic Visualization of a Communal Storytelling Experience”. Atlanta, Georgia: Georgia Institute of Technology.
- Marks, Lawrence E. (1974), “On Associations of Light and Sound: The Mediation of Brightness, Pitch, and Loudness”. *The American Journal of Psychology*, vl. 87, no. 1-2, 173-188.
- Seiça, Mariana (2017), *O Som dos Dados — Sonificação Musical de Emoções Expressas no Twitter*. Dissertação de Mestrado em Design e Multimédia. Coimbra, Portugal: FCTUC.
- Seiça, Mariana, Rui B. Lopes, Pedro Martins e F. Amílcar Cardoso (2017), “Sonifying Twitter’s emotions through music”. *Music Technology with Swing*. Matosinhos, Portugal: Springer, 586-608.
- Schank, Roger C. (1973), *The Fourteen Primitive Actions and Their Inferences*. Stanford University.
- Shwarz, Katharina, Tamara L. Berg e Hendrick P. A. Lensch (2017), “Auto-Illustrating Poems and Songs with Style”. *Computer Vision — ACCV 2016*. Taipei, Taiwan: Springer, 87-103.
- Shultz, Mark F. (2009), “Live Performance, Copyright, and the Future of the Music Business”, *University of Richmond Law Review*, vl. 43, no. 2, 685-764.
- Thayer, R. E. (1989), *The biopsychology of mood and arousal*. Oxford University Press.
- Tschmuck, P. (2012), *Creativity and Innovation in the Music Industry*, 2ª ed., Springer, Viena.
- Tzanetakis, George e Perry Cook (2002), “Musical Genre Classification of Audio Signals”. *IEEE Transactions on Speech and Audio Processing*, vl. 10, no. 5.
- Villegas, Julián e Michael Cohen (2008), ““Roughometer”: Realtime Roughness Calculation and Profiling”. Em: *Proceedings of 125th Audio Engineering Society Convention 2008*, San Francisco, California, October 2-5.
- Yang, Yi-Hsuan e Homer H. Chen (2011), *Music Emotion Recognition*. 1ª ed., CRC Press, Florida, USA.

7.2 Webgrafia

Referências dos links de Internet

Akten, Memo (2008), “Gold”, site disponível em:

<http://www.memo.tv/portfolio/gold/>

(data de consulta: 16.01.2018)

“Almanac: The 45 RPM record”, *CBS News — Sunday Morning* (10 de Janeiro de 2016), site disponível em:

<https://www.cbsnews.com/news/almanac-the-45-rpm-record/>

(data de consulta: 06.01.2018)

Brewer, Jenny (2018), “Brian Roettinger’s album cover for Thirty Seconds To Mars’ America can be customised via online generator”, *It’s Nice That*, (22 de Março), site disponível em:

<https://www.itsnicethat.com/news/willo-perron-thirty-seconds-to-mars-america-album-cover-graphic-design-220318>

(data de consulta: 17.06.2019)

Broderick, Daniel (2014), “9 Album Releases That Turned Music Packaging on Its Head”, *Shutterstock*, (1 de Outubro), site disponível em:

<https://www.shutterstock.com/blog/9-album-releases-that-turned-music-packaging-on-its-head>

(data de consulta: 11.01.2018)

Byrne, Emmet (2016), “Designing Bon Iver’s 22, a Million: An Interview with Eric Timothy Carlson”, *Walker*, (3 de Outubro), site disponível em:

<https://walkerart.org/magazine/designing-bon-iver-22-a-million-album-art-eric-timothy-carlson>

(data de consulta: 17.01.2018)

Coldplay (2015), “Interview: Pilar Zeta (AHFOD artwork creator)”

(20 de Dezembro), site disponível em:

<http://coldplay.com/interview-pilar-zeta-ahfod-artwork-creator/>

(data de consulta: 17.01.2018)

Dredge, Stuart (2013), “Thom Yorke calls Spotify ‘the last desperate fart of a dying corpse’”, *The Guardian*, (7 de Outubro), site disponível em:

<https://www.theguardian.com/technology/2013/oct/07/spotify-thom-yorke-dying-corpse>

(data de consulta: 10.01.2018)

- Doran, John (2009), “Vaughan Oliver: A Portrait Of The Artist As A Jung Man”, *The Quietus*, (30 de Dezembro), site disponível em:
<https://thequietus.com/articles/03454-vaughan-oliver-a-portrait-of-the-artist-as-a-jung-man>
(data de consulta: 17.06.2019)
- Ellis-Petersen, Hannah (2017), “Record sales: vinyl hits 25-year high”, *The Guardian*, (3 de Janeiro), site disponível em:
<https://www.theguardian.com/music/2017/jan/03/record-sales-vinyl-hits-25-year-high-and-outstrips-streaming>
(data de consulta: 16.12.2017)
- Engel, Pamela (2014), “Taylor Swift Explains Why She Left Spotify”, *Business Insider*, (13 de Novembro), site disponível em:
<http://www.businessinsider.com/taylor-swift-explains-why-she-left-spotify-2014-11>
(data de consulta: 09.01.2018)
- Kafka, Peter (2017), “Concert Tours Are Where the Real Money Is”, *ABC News*, (11 de Julho), site disponível em:
<http://abcnews.go.com/Business/story?id=86535&page=1>
(data de consulta: 22.01.2018)
- Kinross, Robin (1990), “Cool, clear, collected”, *Eye Magazine*, site disponível em:
<http://www.eyemagazine.com/%20feature/article/cool-clear-collected>
(data de consulta: 16.06.2019)
- Kissell, Rick (2016), “‘Game of Thrones’ Premiere Draws 8 Million TV Viewers — Just Shy of Series High”, *Variety*, (26 de Abril), site disponível em:
<https://variety.com/2016/tv/news/game-of-thrones-premiere-ratings-season-6-1201759863/>
(data de consulta: 10.06.2019)
- Knopper, Steve (2015), “Islands in the Stream: The 10 Biggest Holdouts in Digital Music”, *The Washington Post*, (2 de Janeiro), site disponível em:
<https://www.rollingstone.com/music/news/artists-refuse-stream-music-20150102>
(data de consulta: 10.01.2018)
- Kreps, Daniel (2008), “Radiohead Publishers Reveal ‘In Rainbows’ Numbers”, *Rolling Stone*, (15 de Outubro), site disponível em:
<https://www.rollingstone.com/music/news/radiohead-publishers-reveal-in-rainbows-numbers-20081015>
(data de consulta: 10.01.2018)

Library of Congress, *Inventing Entertainment: The Early Motion Pictures and Sound Recording of the Edison Companies*: “History of the Cylinder Phonograph”, site disponível em:

<https://www.loc.gov/collections/edison-company-motion-pictures-and-sound-recordings/articles-and-essays/history-of-edison-sound-recordings/history-of-the-cylinder-phonograph/>

(data de consulta: 02.01.2018)

Library of Congress, *Emile Berliner and the Birth of the Recording Industry: “The Gramophone”*, site disponível em:

<https://www.loc.gov/collections/emile-berliner/articles-and-essays/gramophone/>

(data de consulta: 02.01.2018)

Mcintyre, Hugh (2017), “Why Did Taylor Swift Really Rejoin Spotify”, *Forbes*, (27 de Junho), site disponível em:

<https://www.forbes.com/sites/hughmcintyre/2017/06/27/why-did-taylor-swift-really-rejoin-spotify/#7d66446373de>

(data de consulta: 10.01.2018)

Morby, Alice (2017), “Cover for new Beck album Colors is designed to be customised”, *Dezeen*, (30 de Setembro), site disponível em:

<https://www.dezeen.com/2017/09/30/beck-album-artwork-colours-customisable-design/>

(data de consulta: 22.01.2018)

“Os Radiohead já têm a música no Spotify mas Thom Yorke continua zangado”, *BLITZ*, (29 de Dezembro de 2017), site disponível em:

<http://blitz.sapo.pt/principal/update/2017-12-29-Os-Radiohead-ja-tem-a-musica-no-Spotify-mas-Thom-Yorke-continua-zangado>

(data de consulta: 10.01.2018)

Panagiotopoulos, Vas (2015), “The History (and Future) of Live Music”, *NEXT*, (11 de Dezembro), site disponível em:

<https://howwegettonext.com/the-history-and-future-of-live-music-147ecde437b7>

(data de consulta: 23.01.2018)

Richter, Felix (2014), “The LP is Back!”, *Statista*, (6 de Janeiro), site disponível em:

<https://www.statista.com/chart/1465/vinyl-lp-sales-in-the-us/>

(data de consulta: 16.12.2017)

- “Roughness — fluctuation strength”, University of Salford Manchester, site disponível em:
<https://www.salford.ac.uk/research/sirc/research-groups/acoustics/psychoacoustics/sound-quality-making-products-sound-better/accordion/sound-quality-testing/roughness-fluctuation-strength>
(data de consulta: 23.03.2019)
- Shaughnessy, Adrian (1995), “Think of your ears as eyes”, *Eye Magazine*, site disponível em:
<http://eyemagazine.com/feature/article/think-of-your-ears-as-eyes>
(data de consulta: 16.06.2019)
- Siegel, Alan (2015) “‘School’s Out’ Forever: The Secret History Of Alice Cooper’s Classic”, *The Concourse*, (21 de Maio), site disponível em:
<https://theconcourse.deadspin.com/schools-out-forever-the-secret-history-of-alice-coop-1705441582>
(data de consulta: 11.01.2018)
- Temkin, Daniel (2013), “Glitchometry”, site disponível em:
<http://danieltemkin.com/Glitchometry/Index/GlitchometryCircles4>
(data de consulta: 16.01.2018)
- Tyrangiel, Josh (2007), “Radiohead Says: Pay What You Want”, *Time*, (1 de Outubro), site disponível em:
<http://content.time.com/time/arts/article/0,8599,1666973,00.html>
(data de consulta: 10.01.2018)
- “U.S. Sales Database”, RIAA, site disponível em:
<https://www.riaa.com/u-s-sales-database/>
(data de consulta: 06.01.2018)
- Webster, Andrew (2014), “Designer music: how Tycho blends art and sound”, *The Verge*, (25 de Fevereiro), site disponível em:
<https://www.theverge.com/2014/2/25/5419816/tycho-interview>
(data de consulta: 17.01.2018)
- Wilson, Mark (2012), “Synesthesia 2.0: Digital Images, Distorted By Sound Waves”, *Co.Design*, (4 de Novembro), site disponível em:
<https://www.fastcompany.com/1669501/synesthesia-20-digital-images-distorted-by-sound-waves>
(data de consulta: 16.01.2018)

Referências das figuras

Figura 1: Radiohead — *In Rainbows*:

<https://static.stereogum.com/uploads/2017/09/In-Rainbows-1506460221-640x640.jpg>

(data de consulta: 17.06.2019)

Figura 2: The Beatles — *Sgt. Pepper's Lonely Hearts Club Band*:

<https://www2.shutterstock.com/blog/wp-content/uploads/sites/5/2014/10/img251.jpg>

<http://marcmyers.typepad.com/.a/6a00e008dca1f0883401b7c8fd602c970b-500wi>

<https://www2.shutterstock.com/blog/wp-content/uploads/sites/5/2014/10/img250.jpg>

(data de consulta: 17.01.2018)

Figura 3: Alice Cooper — *School's Out*:

https://images.pristineauction.com/29/290162/main_1-Alice-Cooper-Signed-Schools-Out-Record-Album-PA-LOA-PristineAuction.com.jpg

<https://jamesostafford.files.wordpress.com/2013/07/alice-cooper-schools-out-3.jpg>

https://ring.cdandlp.com/gemicricricket/photo_grande/115010343.jpg

(data de consulta: 17.01.2018)

Figura 4: The Rolling Stones — *Sticky Fingers*:

<https://jamesostafford.files.wordpress.com/2013/04/rolling-stones-sticky-fingers-zipper-up.jpg>

<https://d2ydh70d4b5xgv.cloudfront.net/images/e/3/rolling-stones-sticky-fingers-andy-warhol-zipper-cover-coc-59100-bb-aaa-nm-8e4ee8e5098ce4a92781830c3f00e2c8.jpg>

(data de consulta: 17.01.2018)

Figura 5: The Velvet Underground — *The Velvet Underground & Nico*:

<https://www2.shutterstock.com/blog/wp-content/uploads/sites/5/2014/10/img249.jpg>

<https://i.pinimg.com/originals/32/03/83/320383f7273f32639ce257abe241a9cc.jpg>

(data de consulta: 17.01.2018)

Figura 6: Elvis Costello — *Armed Forces*:

https://78.media.tumblr.com/4609b1fec244a1b221b95036c3ef1f53/tumblr_nqn9mcUAEa1rpgpe2o2_1280.jpg

http://www.johncoulthart.com/images/armed_forces/05back4.jpg

http://www.johncoulthart.com/images/armed_forces/O6back5.jpg

(data de consulta: 17.01.2018)

Figura 7: Beck — *The Information*:

<http://www.bigactive.com/wp-content/uploads/2015/04/Beck-CD-Hero-1728x969.jpg>

<http://www.bigactive.com/wp-content/uploads/2015/04/Beck-Wide-Stickers.jpg>

<http://www.bigactive.com/wp-content/uploads/2015/04/Beck-Wide-CDBoxes.jpg>

(data de consulta: 17.01.2018)

Figura 8: Beck — *Colors*:

<http://www.jimmyturrell.com/wp-content/uploads/2017/08/crop-F-Deluxe-overview-v6-1-1024x778.jpg>

http://www.jimmyturrell.com/wp-content/uploads/2017/10/beck-colors-dlx2_1800x-1024x1024.jpg

https://static.dezeen.com/uploads/2017/09/beck-colours-album-cover-jimmy-turrell_dezeen_2364_col_3-1704x852.jpg

(data de consulta: 17.06.2019)

Figura 10: Capas de Reid Miles:

<https://viljamis.com/img/blog/miles/13-e9465723.jpg>

<https://viljamis.com/img/blog/miles/21-a2600e7d.jpg>

<https://viljamis.com/img/blog/miles/98-bf410128.jpg>

<https://viljamis.com/img/blog/miles/15-89b03660.jpg>

<https://viljamis.com/img/blog/miles/12-0f3eb50b.jpg>

<https://viljamis.com/img/blog/miles/10-035ed015.jpg>

(data de consulta: 26.06.2019)

Figura 11: Capas de Robert Flynn:

<http://www.cvinyl.com/images/covers/32165.jpg>

<http://www.cvinyl.com/images/covers/32154.jpg>

<http://www.cvinyl.com/images/covers/32187.jpg>

<http://www.cvinyl.com/images/covers/32143.jpg>

<http://www.cvinyl.com/images/covers/32120.jpg>

<http://www.cvinyl.com/images/covers/32193.jpg>

(data de consulta: 26.06.2019)

Figura 12: Capas de Barbara Wojirsch e Dieter Rehm:

<http://www.hardformat.org/wp-content/uploads/2009/06/bw-1-l.jpg>

<http://www.hardformat.org/wp-content/uploads/2009/06/dr-1-l.jpg>

<http://www.hardformat.org/wp-content/uploads/2009/06/dr-3-l.jpg>

<http://www.hardformat.org/wp-content/uploads/2009/06/bw-4-l.jpg>

<http://www.hardformat.org/wp-content/uploads/2009/06/bw-9-l.jpg>

<http://www.hardformat.org/wp-content/uploads/2009/06/bw-6-l.jpg>

(data de consulta: 27.06.2019)

Figura 13: Capas de Peter Saville:

[https://img.discogs.com/0d7lmyOHYF0SGqYLoq7ms1pi2GA=/fit-in/600x600/filters:strip_icc\(\):format\(jpeg\):mode_rgb\(\):quality\(90\)/discogs-images/R-330199-1513766039-7994.jpeg.jpg](https://img.discogs.com/0d7lmyOHYF0SGqYLoq7ms1pi2GA=/fit-in/600x600/filters:strip_icc():format(jpeg):mode_rgb():quality(90)/discogs-images/R-330199-1513766039-7994.jpeg.jpg)

[https://img.discogs.com/0d7lmyOHYF0SGqYLoq7ms1pi2GA=/fit-in/600x600/filters:strip_icc\(\):format\(jpeg\):mode_rgb\(\):quality\(90\)/discogs-images/R-330199-1513766039-7994.jpeg.jpg](https://img.discogs.com/0d7lmyOHYF0SGqYLoq7ms1pi2GA=/fit-in/600x600/filters:strip_icc():format(jpeg):mode_rgb():quality(90)/discogs-images/R-330199-1513766039-7994.jpeg.jpg)

(data de consulta: 27.06.2019)

Figura 14: Capas de Vaughan Oliver:

https://abcdefghijklmn-pqrstuvwxyz.com/wp-content/uploads/2017/04/O_JoanPons_OliverVaughan_RedHousePainters.jpg

https://abcdefghijklmn-pqrstuvwxyz.com/wp-content/uploads/2017/04/O_JoanPons_OliverVaughan_HisNameIsAlive.jpg

https://abcdefghijklmn-pqrstuvwxyz.com/wp-content/uploads/2017/04/O_JoanPons_OliverVaughan_Breeders.jpg

https://abcdefghijklmn-pqrstuvwxyz.com/wp-content/uploads/2017/04/O_JoanPons_OliverVaughan_ComeOnPilgrim.jpg

(data de consulta: 27.06.2019)

Figura 15: Concerto de Led Zeppelin em Seattle, em 1975:

<https://i.redd.it/Ob80wmgc0wc21.jpg>

(data de consulta: 27.06.2019)

Figura 16: Imogen Heap demonstrando as *Mi.Mu Gloves* na *WIRED 2012*:

https://wi-images.condecn.net/image/R01WdEnbKy8/crop/1620/f/imogen-heap-2_1.jpg

(data de consulta: 27.06.2019)

Figura 18: *The Shape of Story*:

<http://www.durilong.com/research#/shape-of-story/>

(data de consulta: 16.01.2018)

Figura 19: Memo Akten — *Gold*:

<http://www.memo.tv/wpdev/wp-content/uploads/Gold-dust-test-ProRes-048-640x406.jpg>

(data de consulta: 16.01.2018)

Figura 20: Daniel Temkin — *Glitchometry*:

http://danieltemkin.com/Content/Glitchometry/glitch_circles_4.jpg

(data de consulta: 16.01.2018)

<http://danieltemkin.com/Glitchometry/Index/GlitchometryTriangles3>

(data de consulta: 28.06.2019)

Figura 21: FIELD — *Communion*:

https://www.field.io/wp-content/uploads/Projects/field/communion/cache/Communion_1400x0_95.jpg

(data de consulta: 16.01.2018)

Figura 22: The Light Surgeons — *Soundfield*:

<http://www.lightsurgeons.com/site/wp-content/uploads/17.jpg>

(data de consulta: 16.01.2018)

Figura 27: Thirty Seconds To Mars — *America*:

[willo-perron_thirty-seconds-to-mars_album-cover_graphic-design_itsnicethat-names.png](#)

[willo-perron_thirty-seconds-to-mars_album-cover_graphic-design_itsnicethat-apple.png](#)

[willo-perron_thirty-seconds-to-mars_album-cover_graphic-design_itsnicethat-basketball.png](#)

[willo-perron_thirty-seconds-to-mars_album-cover_graphic-design_itsnicethat-3.png](#)

[willo-perron_thirty-seconds-to-mars_album-cover_graphic-design_itsnicethat-4.png](#)

[willo-perron_thirty-seconds-to-mars_album-cover_graphic-design_itsnicethat-pres.png](#)

(data de consulta: 17.06.2019)

Figura 29 e 30: Bon Iver — *22, A Million*:

<http://blogs.walkerart.org/design/files/2016/10/10deathbreastnotes2.jpg>

http://blogs.walkerart.org/design/files/2016/09/BI22_EXC_Poster_02.jpg

http://blogs.walkerart.org/design/files/2016/09/BI22_Gatefold_Outside.jpg

<http://blogs.walkerart.org/design/files/2016/09/PackageComposit.jpg>

http://blogs.walkerart.org/design/files/2016/10/IMG_8454.png

http://blogs.walkerart.org/design/files/2016/10/IMG_8146.png
(data de consulta: 17.01.2018)

Figura 32: Tycho — *Awake*:

<https://serialgk.files.wordpress.com/2014/12/tycho-awake.jpg>
(data de consulta: 17.01.2018)

Figura 33 e 34: Coldplay — *A Head Full Of Dreams*:

<http://coldplay.com/wp-content/uploads/2015/12/p-1.jpg>
(data de consulta: 29.06.2019)
<http://coldplay.com/wp-content/uploads/2015/11/AHFOD1000.jpg>
<http://coldplay.com/wp-content/uploads/2015/12/cpposter.jpg>
<http://coldplay.com/wp-content/uploads/2015/12/e-1.jpg>
<http://coldplay.com/wp-content/uploads/2015/12/wemfinal-1-1.jpg>
(data de consulta: 17.01.2018)

Figura 36: *Modelo W2V*:

<https://github.com/danielvarga/word2vec-web-visualization>
(data de consulta: 14.01.2019)

Figura 37: Árvore da *WordNet*:

https://www.researchgate.net/publication/2841390_Words_with_Attitude/figures?lo=1
(data de consulta: 14.01.2019)

Figura 40: Bailarina:

<https://unsplash.com/photos/pzgnOfeLJwg>
(data de consulta: 23.01.2018)

Figura 45: Imagem recolhida no *Unsplash* para a palavra *sun*:

<https://unsplash.com/photos/ZG6ZepTBHAQ>
(data de consulta: 01.07.2018)

Figura 46 e 49: Imagem recolhida no *Unsplash* para a palavra *balloon*:

<https://unsplash.com/photos/3KEJMfoHoMY>
(data de consulta: 01.07.2018)

Figura 47: Imagem recolhida no *Unsplash* para a palavra *bird*:

<https://unsplash.com/photos/7LsuYqkvIUM>
(data de consulta: 01.07.2018)

Figura 48 e 52: Imagem recolhida no *Unsplash* para a palavra *book*:

https://unsplash.com/photos/GJZXSzjB_AY
(data de consulta: 01.07.2018)

Figura 70: Espectro de cores do *color hue*:

<https://i.stack.imgur.com/YOBFy.png>
(data de consulta: 17.04.2019)

Anexo

Artigo ICCC 2018
Poster ICCC 2018

Anexo

—

Computationally Generating Images for Music Albums

Miguel Cruz, Paul Hardman, João Miguel Cunha

CISUC, Department of Informatics Engineering,
University of Coimbra

migmacruz@gmail.com, {phardman,jmacunha}@dei.uc.pt

Abstract

Nowadays, music albums are seen as a set of unrelated pieces rather than a whole. We propose an approach for a system which produces images for songs of music albums, with the goal of establishing graphic cohesion. It uses semantic analysis of lyrics and semiotic properties to visually represent the meaning and emotions of songs.

Introduction

In the last 20 years there were a lot of changes in the music industry, due to the tremendous technological growth. The new paradigm, dominated by singles and playlists, changed the concept of music album from something which was normally seen as a whole into a simple set of individual and unrelated pieces. Moreover, the relation between the music album and its related concerts has been reduced – more importance is given to festivals with several music artists rather than to individual concerts (Tschmuck 2012).

The main goal of this project is to use computational means to integrate all the materials and events related to a music album, making it work both individually and as a whole. In order to achieve this, we propose a computational approach that produces images from the song lyrics of an album and transforms them according to sound variables of the songs. These images can then be used to (i) create the graphic materials of the album, (ii) produce videos for the songs and (iii) visual effects for the corresponding concerts. This way, all of the materials and events will be integrated in one single graphic system, fully related to the music album.

Our Approach

The relation between text, image and sound serves as a basis to the proposed computational system, which is divided into six layers (L1-L6). This modular architecture allows an iterative development towards our final goal: having images that illustrate the songs' lyrics and are distorted according to the sound of the voice and instruments, using semiotic properties to visually represent the emotions of the songs.

L1 Lyrics Analysis The first layer is responsible for the analysis of the lyrics, which retrieves words that can be searched in an image database. In order to do this, the system uses the Processing RiTa library to divide the lyrics in nouns, adjectives and verbs.

L2 Lyrics Preparation In this layer, the user is asked to determine the time in the song in which each retrieved word appears, in order to know when the images should be created or distorted. This achieves synchronization between song and image.

L3 Image Retrieval The nouns retrieved by L2 are used in a search process conducted in an image database (e.g. Unsplash), gathering images that illustrate them. Our goal is to follow a semi-automatic approach, in which the user has the possibility of selecting which gathered images should be used.

L4 Image Preparation This layer uses the images provided by the previous layer and applies a filter to them, making them better suited for visual transformations. Currently, we are using OpenCV Canny Edge Detector to transform the images into a set of ellipses. We plan to explore other possibilities in the future: e.g. line representation or even clustering-techniques.

L5 Illustration Production The graphic elements, produced by the previous layer from the gathered images, are used in combination with semantic analysis to produce illustrations. This process uses the verbs and adjectives retrieved using RiTa and uses them to apply visual transformations, based on semantic-semiotic mappings (e.g. words related to movement affect element positioning).

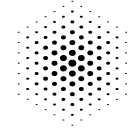
L6 Background Variation The background of the illustration will be affected by sound variables of the instrumental. One example is changing the colour according to the musical scales or chords that are being played at each moment: colder colours if the scales/chords are minor and warmer colours if they are major. In order to do this, we will orientate the study to the piano to make a MIDI analysis in Pure Data. The program could have another component consisting in an estimated analysis of the full instrumental.

Future Work Currently, the user is manually searching the images, as the image retrieval is not yet implemented. In the future we intend to explore with other filters for the Image Preparation layer.

References

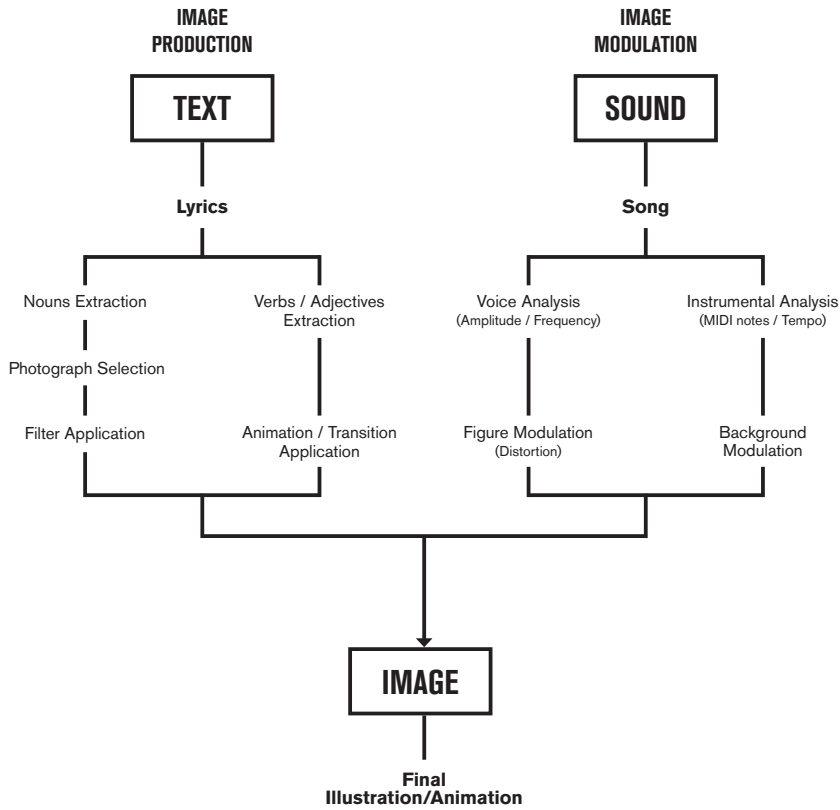
Tschmuck, P. 2012. Creativity and innovation in the music industry. In *Creativity and Innovation in the Music Industry*. Springer. 225–251.

COMPUTATIONALLY GENERATING IMAGES FOR MUSIC ALBUMS



COMPUTATIONAL
DESIGN &
VISUALIZATION
LAB.

MIGUEL CRUZ • MIGMCRUZ@GMAIL.COM / PAUL HARDMAN • PHARDMAN@DEI.UC.PT / JOÃO M. CUNHA • JMACUNHA@DEI.UC.PT



Selected Photo for the noun "dance"



Canny Edge Detector Filter



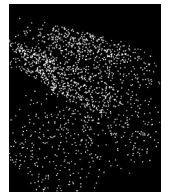
Illustration Production



Output Illustration



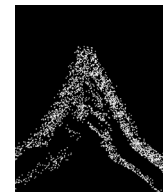
Animated Illustration



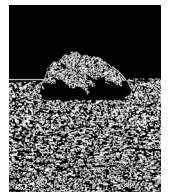
Animated Transition



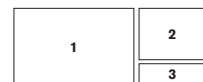
Bird



Hands



Rock



1. Structure of the system for the image production and modulation. Given the lyrics of a song, the programme extracts the nouns, verbs and adjectives (L1). The nouns serve as keywords to the photograph search (L3). After the selection, the Canny Edge Detector Filter is applied (L4) and then the illustration is produced over its output. In this step, the verbs/adjectives that appear will influence the animation of the illustration or the transition between two different illustrations. At the same time, the sound of the voice is being analysed to modulate the figure (L5). The instrumental is also being analysed to affect the background of the illustration (L6). All this path will output several illustrations that can be used afterwards.

2. An example of some steps of the system applied to an image retrieved from the noun "dance".

3. Some other examples of output illustrations from different nouns.

In the last 20 years there were a lot of changes in the music industry, due to the tremendous technological growth. The new paradigm, dominated by singles and playlists, changed the concept of music album from something which was normally seen as a whole into a simple set of individual and unrelated pieces. Moreover, the relation between the music album and its related concerts has been reduced — more importance is given to festivals with several music artists rather than to individual concerts (Tschmuck 2012).

The main goal of this project is to use computational means to integrate all the materials and events related to a music album, making it work both individually and as a whole. In order to achieve this, we propose a computational approach that produces images from the song lyrics of an album and transforms them according to sound variables of the songs. These images can then be used to (i) create the graphic materials of the album, (ii) produce videos for the songs and (iii) visual effects for the corresponding concerts. This way, all of the materials and events will be integrated in one single graphic system, fully related to the music album.

Our Approach

The relation between text, image and sound serves as a basis to the proposed computational system, which is divided into six layers (L1-L6). This modular architecture allows an iterative development towards our final goal: having images that illustrate the songs' lyrics and are distorted according to the sound of the voice and instruments, using semiotic properties to visually represent the emotions of the songs.

L1: Lyrics Analysis. The first layer is responsible for the analysis of the lyrics, which retrieves words that

can be searched in an image database. In order to do this, the system uses the Processing RiTa library to divide the lyrics in nouns, adjectives and verbs. The nouns will work as keywords for the search in the image database (as explained in L3). The verbs will be mapped into basic actions that have associated animations. This component was inspired by Duri Long's *The Shape of Story*, in which the verbs appearing in the storytelling are mapped into fourteen basic actions, adapted from the work of Roger C. Shank: smell, move, think, ingest, speak, see, hear, feel, have, conclude, be, transport, expel, propel (Long 2017). We intend to create groups for these actions, mapping them into five or six animations, instead of fourteen. The adjectives will affect other properties of the image or the animation.

L2: Lyrics Preparation. In this layer, the user is asked to determine the time in the song in which each retrieved word appears, in order to know when the images should be created or distorted. This achieves synchronization between song and image.

L3: Image Retrieval. The nouns retrieved by L2 are used in a search process conducted in an image database (e.g. Unsplash), gathering images that illustrate them. Our goal is to follow a semi-automatic approach, in which the user has the possibility of selecting which gathered images should be used.

L4: Image Preparation. This layer uses the images provided by the previous layer and applies a filter to them, making them better suited for visual transformations. Currently, we are using OpenCV Canny Edge Detector to transform the images into a set of ellipses. We plan to explore other

possibilities in the future: e.g. line representation or even clustering-techniques.

L5: Illustration Production. The graphic elements, produced by the previous layer from the gathered images, are used in combination with semantic analysis to produce illustrations. This process uses the verbs and adjectives retrieved using RiTa and uses them to apply visual transformations, based on semantic-semiotic mappings (e.g. words related to movement affect element positioning).

L6: Background Variation. The background of the illustration will be affected by sound variables of the instrumental. One example is changing the colour according to the musical scales or chords that are being played at each moment: colder colours if the scales/chords are minor and warmer colours if they are major. In order to do this, we will orientate the study to the piano to make a MIDI analysis in Pure Data. The program could have another component consisting in an estimated analysis of the full instrumental.

Future Work

Currently, the user is manually searching the images, as the image retrieval is not yet implemented. In the future we intend to explore with other filters for the Image Preparation layer.

References

- [1] Tschmuck, P. 2012. Creativity and innovation in the music industry. In *Creativity and Innovation in the Music Industry*. Springer. 225–251.
- [2] Long, D. et al. 2017. *The Shape of Story: A Semiotic Artistic Visualization of a Communal Storytelling Experience*.

