U.PORTO

MESTRADO
MULTIMÉDIA - ESPECIALIZAÇÃO EM TECNOLOGIAS

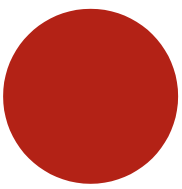
AUDIÇÃO MUSICAL *AFETIVA*: RECURSO A EXPRESSÕES FACIAIS PARA RECOMENDAÇÃO MUSICAL

João Pedro Dias Babo de Carvalho

M
2018

FACULDADES PARTICIPANTES:

FACULDADE DE ENGENHARIA
FACULDADE DE BELAS ARTES
FACULDADE DE CIÊNCIAS
FACULDADE DE ECONOMIA
FACULDADE DE LETRAS



U.PORTO

Audição Musical *Afetiva*: Recurso a expressões faciais para recomendação musical

João Pedro Dias Babo de Carvalho

Mestrado em Multimédia da Universidade do Porto

Orientador: Doutor Matthew Edward Price Davies (Investigador Sénior)

Coorientador: Mestre Luciano José Santos Reis Moreira (Assistente Convidado)

Junho de 2018

Audição Musical *Afetiva*: Recurso a expressões faciais para recomendação musical

João Pedro Dias Babo de Carvalho

Mestrado em Multimédia da Universidade do Porto

Aprovado em provas públicas pelo Júri:

Presidente: Professor Hugo José Sereno Lopes Ferreira (Professor Auxiliar)

Vogal Externo: Doutora Ana Maria Silva Rebelo (Investigadora Sénior)

Orientador: Doutor Matthew Edward Price Davies (Investigador Sénior)

Resumo

É uma tarefa árdua democratizar a música em função dos gostos pessoais de cada indivíduo. Os sistemas de recomendação de música baseiam genericamente as suas propostas em similaridades musicais, segundo a análise do conteúdo áudio ou um *feedback* de preferência, não contemplando as reações dos estados emocionais dos ouvintes. Este estudo explora a possibilidade de tornar a recomendação de música numa relação mais "pessoal" entre os ouvintes e o sistema, registando as suas respostas emocionais. Assim, investigamos se o reconhecimento automático das emoções com recurso a expressões faciais poderá servir de veículo para uma escolha musical personalizada. Um conjunto de 26 participantes foi exposto à audição de 17 excertos musicais cujas partes continham o verso e o refrão do tema. Foi usado um *Software Development Kit* que permitiu, com o auxílio de uma *webcam* registar as expressões faciais de cada participante durante a audição. Obtiveram-se dados tanto sobre as emoções como as expressões faciais presentes ao longo do tempo em cada amostra musical apresentada. Foi recolhida uma avaliação por intermédio de um questionário, cuja parte principal se baseava numa escala de tipo *Likert* de sete pontos sobre o prazer associado à audição de cada tema por cada participante. Na análise dos dados recolhidos, três correlações significativas foram encontradas que relacionam duas expressões faciais e um estado emocional com as médias de avaliação das músicas de todo o grupo de participantes. Mais correlações significativas podem ser encontradas analisando os participantes individualmente sugerindo que as expressões faciais são indicadores subtis de preferência musical.

Fica em aberto o caminho para uma interpretação individual num contexto de *machine learning* da vontade explícita de cada um de mudar de música.

Palavras-chave: recomendação de música; expressões faciais afetivas

Abstract

It’s a hard task to democratize music regarding listeners’ personal preferences. Music recommendation systems generally base their proposals in music similarities according to audio content analysis or feedback ratings about preference, and typically ignore the emotional context of the listeners. This study explores the possibility of making music recommendation more “personal” relationship by capturing the emotional response of the listeners. This dissertation seeks to determine whether the automatic recognition of emotions using listeners’ facial expressions could drive the personalized recommendation of music. A group of 26 participants listened to 17 musical excerpts composed by verse and chorus. A state of the art facial emotion detection SDK (Software Development Kit) with the aid of a webcam was used to register the facial expressions of each of the participants while they listened to the musical excerpts. Data was extracted about the presence of emotions and facial expressions during the listening time of each presented song. In addition, each listener rated their opinion about each musical excerpt on a Likert scale. Based on the analysis of the acquired data, three significant correlations could be found which related two facial expressions and one emotional state to the mean ratings of the entire group of participants. Further to this, many significant correlations could be found when analyzing the participants independently. Thus, suggesting that facial expressions are subtle indicators of musical preference, but the most informative facial cues typically vary according to individuals.

It’s revealed the path to an individualized interpretation in a machine learning context of each ones explicit will of change the playing music.

Keywords: music recommendation; affective facial expressions

Agradecimentos

O meu mais sincero e profundo agradecimento ao meu elo de ligação ao mundo… a minha Mãe… pelo amor, pelo apoio, pela paciência… A pessoa mais importante, sempre. Ao meu Pai que o sinto sempre comigo. Ao José Alberto Alves pela amizade, pelo apoio, pela curiosidade e o sorriso ao final da noite.

Ao meu orientador Dr. Matthew Davies e ao meu coorientador Prof. Luciano Moreira pelo voto de confiança, pela simpatia, pelo profissionalismo, pelo entusiasmo nas nossas reuniões que me deu ânimo quando tudo parecia estar a desmoronar-se.

Ao João Ferreira, amigo de longa data, ao Johnny Murata e ao Pedro pelo incentivo à candidatura para o Mestrado Multimédia.

Ao Prof. Dr. Rui Penha e ao Eduardo Magalhães pelo sentido crítico e disponibilização do Laboratório Áudio para realização dos testes.

Aos *Fantásticos Multimédicos*, Gaspar Lopes, Guida Rocha, Valter Abreu, Ricardo Carvalho, Tuca Silveira, Alonso Torres, Francisco *Chico*, Sara Eleanor Barreira, Pedro Ayala, Ana Rita Costa, João Paiva e aos restantes alunos com quem me cruzei, pela amizade, brincadeira e *trocas acesas* de opinião.

Ao Eng. Rogério Soares pela ajuda na programação, pelos momentos musicais partilhados e pela crítica sempre incisiva.

À Psicóloga Ana Ribeiro que prontamente acedeu aos meus pedidos de opinião.

Aos meus amigos e colegas músicos que me solicitaram um infindável número de vezes a explicação daquilo *que eu estava na realidade a fazer…*

Ao Sound and Music Computing Group do INESC Porto.

À Cátia Faria, *A Filósofa*, por me servir de referência e inspiração e a todos os participantes na experiência realizada neste trabalho.

Muito obrigado!

João Pedro Dias Babo de Carvalho

Índice

1. Introdução

A música está hoje em dia presente no nosso meio ambiente numa escala nunca antes vista. A tecnologia e sua portabilidade permitem que tenhamos como companhia os nossos artistas preferidos e que possamos de forma simples descobrir outros focos de interesse musical. Seja para nos sentirmos melhor, para prolongar um estado emotivo, ou um ato puramente lúdico, existem opções tecnológicas de recomendação musical que pretendem satisfazer os nossos intentos.

Este trabalho baseia-se no estudo das expressões faciais veiculadas pelas emoções no que respeita à audição musical *afetiva*. Este termo hoje em dia descreve a ideia de uma relação emocional perante determinado objeto, sendo nesta abordagem a música.

No entanto, os sistemas de recomendação musical não contemplam o estado nem as reações emotivas do ouvinte veiculadas por expressões faciais, ignorando a altura de mudar o registo musical ou eventualmente insistir numa aproximação ao registo anterior. A análise de expressões faciais pode ajudar no processo de perceber em que momento o ouvinte deseja mudar de música resultando em dados suscetíveis de alimentar um sistema automático de recomendação de musical. Explora-se a possibilidade de recomendar de forma mais pessoal e a identificação de expressões faciais e emoções que se relacionem com o prazer na audição de música.

É possível detetar de forma automática a vontade de mudar? Será o estado emocional o melhor indicador para alimentar um sistema de recomendação automática de música?

Com recurso a um *software* de análise de expressões faciais e emoções e uma *webcam*, 26 participantes ouviram 17 excertos musicais cada um, atribuindo-lhes no final uma avaliação quantitativa respeitante às suas preferências musicais.

No capítulo 2 são referenciadas diferentes abordagens de recolha de dados para recomendação musical e as suas fragilidades. No capítulo 3 é descrito o método de recolha

de dados nesta dissertação. Os capítulos 4 e 5 apresentam respetivamente os resultados obtidos e a sua discussão.

O último capítulo é uma reflexão sobre o trabalho efetuado dando resposta às questões colocadas preconizando explorações futuras no domínio do uso das expressões faciais como modelo individual de caracterização dos ouvintes servindo de adjuvante à evolução de modelos de sistemas de recomendação automática de música.

Este trabalho de caráter exploratório contribui com uma avaliação de quais os melhores indicadores emocionais e expressões faciais do prazer na audição musical que podem suportar a vontade de mudar de música e descreve um método de recolha, organização e avaliação dos dados.

2. Estado da Arte

2.1. Introdução

Este capítulo contextualiza historicamente o estudo das emoções e expressões faciais do ser humano. Faz referência a modelos de estudo para caraterização das emoções e a sistemas de recomendação de música apresentando as suas fragilidades.

2.2. Expressões faciais *afetivas*

A cara além de nos identificar e distinguir, é um dos mais importantes canais de comunicação não verbal de que dispomos.

As expressões faciais podem representar os nossos estados *afetivos*, isto é, relativos à emoção, ou ser o veículo de informação não emocional. Podem demonstrar uma intenção, comunicar sinais culturais específicos (por ex. piscar o olho), e podem indicar estados clínicos de doença ou dor (Baltrusaitis, 2014).

O estudo das expressões faciais relativas à emoção iniciou-se no séc. XIX por um neurologista francês Duchenne de Boulogne que tentou identificar os músculos faciais envolvidos nalgumas expressões tais como o sorriso associado à alegria e o sorriso como manifestação de outras emoções (Keltner, Ekman, Gonzaga, & Beer, 2003).

Duas abordagens acerca das emoções foram propostas. Darwin considerou que as emoções são modulares ou discretas e usou terminologia para as diferenciar. Termos como raiva (*anger*), medo (*fear*), repugnância (*disgust*) entre outros, foram definições encontradas como modo de distinção entre diversos estados emotivos no ser humano. Em

1896, Wundt propôs que as emoções fossem interpretadas em função das dimensões agradável e desagradável e alta e baixa intensidade. Mais tarde considerou também as emoções teriam um carácter modular: cada emoção seria descrita como uma combinação dos métodos modular e dimensional. Cada módulo seria interpretado em função da variação das suas dimensões constituindo o modelo conhecido como *Valence-Arousal* (Ekman, 2016).

Apoiado no trabalho de Duchenne, Darwin popularizou-o publicando fotografias de expressões faciais usando-as para perceber se a avaliação feita pelos participantes nessa experiência correspondia à atribuição da mesma emoção (Baltrusaitis, 2014) mostrando evidências da sua universalidade cultural (Keltner et al., 2003).

Paul Ekman, no séc. XX, deu continuidade à investigação das emoções concluindo que o ser humano tem a habilidade inata de expressar e interpretar seis consideradas universais; *happiness*, *anger*, *disgust*, *fear*, *surprise* e *sadness* estando a sua causa e intensidade dependentes de fatores culturais (Ekman & Friesen, 1975). No entanto, as mesmas expressões faciais avaliadas em culturas distintas são descritas como relativas à mesma emoção (Paul Ekman, 2004). As emoções poderiam, de facto, ser um produto da cultura e consequentemente não corresponder à emoção sentida, mas em resposta a estímulos semelhantes em ambiente privado verificou-se que americanos e japoneses apresentavam as mesmas expressões. Quando confrontados com situações de rotina do dia a dia os japoneses mascaravam mais as emoções negativas com um sorriso do que os americanos (Paul Ekman, 2004). Partilhamos alguns processos que estimulam as mesmas reações emocionais tal como partilhamos as expressões para cada emoção, mas há processos que despoletam emoções que não são só específicos de uma cultura, são individuais (Paul Ekman, 2004).

Encontradas as evidências de expressões faciais da emoção universais, Paul Ekman e Wallace Friesen investigaram qual seria o aspeto físico das expressões desenvolvendo uma tabela representativa de cada músculo da cara. Nesta tabela foram descritos os músculos envolvidos em cada expressão facial das emoções consideradas como primárias e universais (P Ekman & Friesen, 1975). A tabela tem a designação de *Facial Action Coding System (FACS)* atribuindo um número a cada músculo facial que

se designa por *Action Unit (AU)*. Cada emoção das 6 referenciadas é composta por um conjunto de *Action Units*, ou seja, um conjunto de movimentos de alguns músculos que caracterizam a expressão facial correspondente (Cohn, Ambadar, & Ekman, 2007).

Estudos mais recentes sugerem só quatro emoções como sendo básicas em vez das seis descritas (JACK et al., 2014). A análise de cada musculo facial ativado na representação de emoções demonstrou diferenças claras entre as expressões de *happiness* e *sadness* mas algumas semelhanças foram encontradas entre *fear*, e *surprise*, *anger* e *disgust*. No entanto existe hoje em dia um consenso pela comunidade científica que reconhece uma universalidade às emoções consideradas básicas. Um questionário feito a 248 cientistas em junho de 2014 revela que 91% das respostas indica que concordam que a emoção anger é uma emoção básica, 90% fear, 86% disgust, 80% sadness, 76% happiness e surprise obteve 50 (Paul Ekman, 2016).

A espécie humana revela mais expressões associadas à emoção que são reflexo da combinação das chamadas emoções básicas. Por exemplo *happily surprised e angrily surprised* combinam movimentos musculares descritos na tabela *Facial Action Coding System* observados nos estados de *happiness* e *surprised* e, *angry* e *surprised* (Du, Tao, & Martinez, 2014). Este tipo de emoções é descrito pelo modelo *Valence-Arousal*.

2.3. Sistemas de recomendação de música

Uma área que tem sido negligenciada na recomendação automática de música é o estado e reações emocionais do ouvinte. As abordagens mais comuns permitem fazer recomendações a partir das mais diversas características e categorias atribuídas às músicas.

Os componentes comuns de um sistema de recomendação são três e relacionam-se com o perfil do utilizador, o item ou artigo objeto de consumo e o conjunto de algoritmos que estabelece uma relação entre o utilizador e o item.

Criar um perfil de utilizador é o processo de identificar os dados acerca do interesse do utilizador num determinado domínio (Kanoje, Girase, & Mukhopadhyay, 2014).

O perfil do utilizador descreve os seus gostos musicais e divide-se em duas etapas: *User Profile Modelling* e *User Listening Experience Modelling*. A primeira contém informação (Celma Herrada., 2009) que descreve a idade, o género, o país e a cidade, previsões a longo prazo tais como interesses e personalidade e atributos que podem variar de hora em hora tal como o humor, atitude, etc. Os dados relativos ao utilizador podem ser obtidos por um processo explícito que questiona diretamente o utilizador para obtenção de dados ou implícito que avalia as ações e comportamentos do utilizador (Kanoje et al., 2014).

A segunda utiliza uma taxonomia constituída por quatro categorias baseadas num estudo que analisou diferentes tipos de utilizadores cuja idade varia entre 16 e 45 anos: *savants, enthusiasts, casuals* e *indifferents* (D Jennings,2007).

O perfil do item, neste caso música, classifica-a em função de três categorias: editorial, cultural e acústica. A editorial pode conter o compositor e o título; a cultural é o resultado da análise de informação relativa à comparação de músicas semelhantes e por fim a acústica é relativa à analise dos sinais de áudio (tonalidade, compasso, etc.) (Song, Dixon, & Pearce, 2012).

O relacionamento dos dois perfis, o do utilizador com o da música divide-se em seis tipos comuns de abordagem: *metadata information retrieval, collaborative filtering, content-based music information retrieval, emotion-based model, context-based information retrieval* e *hybrid models*.

O método de *metadata information retrieval* usa informação do tipo editorial fornecida pelos criadores e /ou editores como por exemplo o título da música e o nome do artista (Downie, 2005). Trata-se de um modelo em que o utilizador tem que conhecer a existência de dados para uma música em particular. Torna-se também fastidioso manter a base de dados e não é considerada nenhuma informação acerca do utilizador.

O método *Collaborative Filtering* reúne três tipos de abordagens. A *Memory-based Collaborative Filtering* recomenda baseado em avaliações prévias feitas pelos utilizadores agrupando-os segundo interesses semelhantes de forma a que um novo item seja calculado usando um grande número de votos explícitos que estejam fortemente correlacionados. Contrastando com o método anterior, o *Model-based Collaborative Filtering* usa *machine learning* e algoritmos de *data mining* que permitem ao sistema treinar um modelo das preferências dos utilizadores que contém uma lista de valores das avaliações feitas e constrói um modelo de previsão do próximo item. A terceira abordagem é um *Hybrid Collaborative Filtering* que combina diversos modelos de *Collaborative Filterings*. Está provado que uma abordagem híbrida tem uma melhor performance do que qualquer um dos modelos de forma individual (Wang, de Vries, & Reinders, 2006). Uma limitação é o facto de a tendência de avaliações mais elevadas ter como consequência que as músicas menos cotadas não sejam tão visíveis para os utilizadores (Herlocker, Konstan, Terveen, & Riedl, 2004).

A técnica de *Content Music Information Retrieval* recomenda em função de análises feitas ao áudio das músicas já ouvidas em vez de se basear nas avaliações feitas pelo utilizador. Algumas das características mais representativas são o timbre e o ritmo (Cano, Koppenberger, & Wack, 2005). A distância entre músicas, ou seja, elementos correlacionados que lhes conferem alguma semelhança, é calculada tipicamente por três processos. O primeiro gera uma distância entre uma mistura de modelos Gaussianos combinando distâncias individuais entre componentes gaussianos (Salomon & Logan, 2001). O segundo usa vetores cuja amostra é retirada do cálculo *Gaussian Mixture Model* de duas músicas a ser comparadas. A amostragem é feita por geração aleatória de números (Elias Pampalk Tim, Tim Pohle, 2005). O terceiro calcula estatísticas como a média e a variância (Chordia, Godfrey, & Rae, 2008). Medindo distâncias entre semelhanças de características acústicas esta abordagem resolve alguns problemas. No entanto comportamentos semelhantes dos utilizadores podem não levar à escolha das mesmas músicas.

A abordagem ao *Emotion-based Model* é uma tendência que relaciona o contexto emocional dos utilizadores com a música (Kim et al., 2010). Relaciona o modelo *valence-*

arousal que define a emoção percecionada pelo utilizador e a associa a características acústicas da música. Uma limitação para modelar o sistema é a necessidade de uma grande quantidade de dados. Existe também o problema da interpretação individual de cada utilizador e a forma como emotivamente se expressa. Seguindo um caminho bastante diferente o *Context-based Information Retrieval* não usa informação acústica nem avaliações dos utilizadores, centrando-se na opinião pública (dos utilizadores) para identificar e recomendar música.

Usa técnicas para filtrar informação sobre artistas semelhantes, classificação do estilo musical, dados sobre a emoção, etc.

Como consequência, a música mais popular pode ter mais avaliações e ofuscar outros artistas havendo um problema de tendências de mercado e publicidade.

As abordagens híbridas (*Hybrid Model Information Retrieval*) combinam mais do que um modelo numa tentativa de melhorar o serviço de recomendação.

Há no mercado três sistemas comerciais de recomendação incontornáveis: *Pandora, Spotify* e *iTunes Genius*.

O *Pandora Internet Radio* (*Pandora*) é um exemplo comercial de um sistema de recomendação. Deste sistema faz parte o projeto *"Music Genome Project"* que modelou músicas com vetores definidos por cerca de 450 diferentes características sendo cada uma delas atribuída por "musicólogos profissionais" (Barrington et al., 2009). Este método usado pelo *Pandora* é fastidioso considerando o tempo despendido para classificar os atributos de centenas de milhares de músicas. Existe neste processo o problema da introdução de ambiguidade de critérios pois na perceção musical, uma vez que música de culturas diferentes provoca diferentes respostas em função da cultura musical do avaliador (Dalton et al., 2016). O *Music Genome Project* é um exemplo de um algoritmo com filtro do tipo *content-based* (Bogdanov 2011), em que uma música é descrita usando as suas características intrínsecas em vez de dados tais como o nome do artista, nome do álbum, género de música, etc.

O *iTunes Genius* usa, em contraste ao *Music Gnome project*, um filtro colaborativo sendo as músicas recomendadas em função da compra de canções dos outros utilizadores (Cremonesi, Garzotto, Negro, Papadopoulos, & Turrin, 2011).

Este sistema com uma base de dados massiva de tendências de compra de músicas fornece informação ao serviço de recomendação do *iTunes*, o *ITunes Genius* (Barrington et al., 2009). O *iTunes usa o Gracenote´s MusicID* para reconhecer músicas, artistas e álbuns da biblioteca de um utilizador. Essa informação combinada com a classificação dada pelo utilizador às canções é comparada com os metadados que o *Genius* possui para gerar *playlists* da biblioteca disponível e recomenda novas compras (Barrington et al., 2009). Este processo assenta nas tendências de compra de música gerais dos utilizadores ignorando as preferências pessoais de cada um. Este tipo de filtro colaborativo, no entanto, relega para segundo plano artistas menos conhecidos que estão fora das tendências atuais de compra sendo muito menor a possibilidade da sua recomendação (Magno & Sable, 2008; Barrington, Oda, & Lanckriet, 2009).

O *Spotify* evoluiu de um site de *music streaming* para um sistema de recomendação específico para cada utilizador após ter adquirido a plataforma *Echo Nest* em 2014. Esta plataforma utiliza métodos de processamento de sinal que permitem o uso por outras companhias através da *API's* (*Application program interface*). O *Spotify* serve-se dos dados do *Echo Nest* que contém atributos como o tempo da música (bpm), a tonalidade, o compasso, valores sobre a energia e *loudness*, dados sobre o artista tais como a sua biografia, notícias e artistas semelhantes ("*Spotify Echo Nest API*"). Estes dados quantitativos são combinados de uma forma não revelada pelo *Spotify*, mas que geram recomendações e playlists personalizadas para os seus utilizadores. Este processo de abordagem híbrida não integra quaisquer informações acerca do contexto do ouvinte nem a sua apreciação emocional acerca da música (Dalton et al., 2016).

Em 2016 uma abordagem híbrida de recomendação de música foi testada pelo *Computer Science and Engineering Department, Chandigarh University, Mohali* na Índia. O objetivo é recomendar música ao ouvinte que se enquadre nos estilos musicais que o ouvinte normalmente escolhe (Singh & Boparai, 2016). Porém, o sistema não considera a reação emocional do utilizador. Baseia-se em semelhanças dos tipos de utilizadores e usa algoritmos para encontrar as músicas mais ouvidas e criar listas de recomendação.

2.4. Emoções e computação *afetiva*

Desde o início da humanidade que a música tem tido um papel relevante na sociedade tal como a conhecemos. Independentemente do contexto cultural, social e até de localização geográfica, a música é inserida nas vivências diárias de cada um em manifestações como casamentos, funerais, para praticar desporto, trabalhar, etc.

Investigadores no campo da cognição e da neurociência, que afirmam que as emoções são uma das principais razões pela qual se ouve música, advogam que a música pode expressar e induzir emoções. Quando o ouvinte associa alguma emoção a uma passagem ou obra musical, trata-se, de acordo com tais investigadores, de uma expressão da própria música (por exemplo: esta peça é triste). Considera-se que as emoções são induzidas no ouvinte quando se sente por exemplo triste durante ou depois da obra musical ter acabado. (Juslin and Sloboda 2011)

Definir os conceitos de *mood*, *emotion*, *affect* e *feeling* distintos teoricamente, mas muitas vezes aplicados de uma forma menos correta (trocar *mood* por *emotion*, p.ex.) é um desafio. (Russell & Barrett, 1999; Scherer, 2005). Uma perspetiva aceite é a de que a emoção é constituída por: *cognitive appraisal*, *bodily reactions*, *action tendencies*, *motor expressions*, e *subjective feelings* (Scherer, 2005). Segundo esta perspetiva o termo *feeling* é muitas vezes interpretado como sinónimo de *emotion*. *Moods*, são difusos e ocorrem durante *affective states*, tendo uma intensidade menor que as emoções, duram menos tempo e não têm um objeto claro. O termo "*affect*" surge como um unificador descritivo englobando *emotion*, *mood* e também *preferences* (Pasi, 2015).

Muitas vezes os termos *emotion*, *mental state* e *affective* são usados no mesmo contexto referindo-se a um estado dinâmico derivado da experienciação de um sentimento (Baltrusaitis, 2014).

Há hoje em dia a convicção que os sistemas informatizados com a capacidade de detetarem affective states dos utilizadores é benéfica. Na área de estudos de *Affective Computing* tenta-se colmatar a diferença entre a expressividade emocional humana e o computador desprovido de emoções.

Russell and Barrett (1999) fizeram a distinção entre *prototypical emotions* e "*core affects*". As *protoypical emotions* têm um objeto às quais são dirigidas tal como pessoas, *conditions or events*, enquanto as *core affects* não são necessariamente dirigidas a nada. Uma *mood* foi também definida por eles como um *prolonged core affect*.

As emoções envolvidas na audição de música são diferentes das emoções do dia a dia (Juslin, 2013ª). Os *affective states* provocados pela audição de música são aproximados às emoções experienciadas no dia a dia tais como *happiness* e *sadness* mas também se relacionam com emoções estéticas ou emoções específicas da música induzidas pela apreciação das suas qualidades intrínsecas (Scherer, 2004; Juslin, 2013ª), tais como *wonder*, *admiration* e *solemnity*. Pode também haver uma mistura de emoções envolvida na audição de música tal como *happiness* e *sadness* (Gabrielsson 2010).

Quando se estuda emoções no contexto musical há que fazer uma distinção entre emoções provocadas no ouvinte como consequência da audição musical e as emoções que o tema musical exprime.

Sabe-se que as emoções induzidas pela música não são necessariamente as mesmas que a peça musical possa ter como pretensão (Gabrielson, 2002). As emoções expressas pelo ouvinte são únicas, podem variar de individuo para individuo (Juslin, 2013b).

Apesar de do ponto de vista da MIR (*Music Information Retrieval*), a *mood* estar relacionada com características musicais inerentes à própria peça, em última análise a *mood* de uma música é composta pela avaliação individual de cada ouvinte (Juslin, 2013b). As *moods* percecionadas são influenciadas por vários fatores: a estrutura do tema musical (Gabrielsson & Lindstrom, 2001), características individuais da pessoa tais como a personalidade (Vuoskoski & Eerola, 2011), e o contexto da audição (Scherer & Zentner, 2001). Apesar da influência dos fatores relacionados com o ouvinte e o contexto da audição limitarem a *mood* que possa advir da estrutura de uma peça musical, a estrutura transcende a cultura dos ouvintes (Balkwill & Thompson, 1999; Fritz et al.,2009) e não é dependente do conhecimento musical de cada individuo (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005) (Pasi, 2015). As alterações na fisiologia das emoções estão associadas as expressões reproduzidas facialmente (Keltner et al., 2003). Medições

de EEG da atividade cerebral mostraram que sujeitando indivíduos a estímulos de carácter emotivo algumas assimetrias surgiram quando as emoções eram negativas ou positivas.

Pesquisas recentes com vídeos que induzem emoções em que foram gravadas as respostas fisiológicas e expressões faciais dos indivíduos participantes com eletroencefalograma(EEG) e com software de reconhecimento de expressões faciais, demonstram que os resultados das expressões são superiores aos do EEG (Soleymani, Asghari-Esfeden, Fu, & Pantic, 2016).

3. Método

3.1. Resumo

A experiência da audição de música é suscetível de criar estados emocionais nas pessoas. Os sistemas de recomendação genericamente valorizam a similaridade musical não contemplando o estado emocional dos ouvintes e o modo como este influencia as suas preferências musicais num determinado momento. Numa perspetiva de perceber a possibilidade de prever a adequação de determinada música ao desejo do ouvinte num determinado momento, este capítulo descreve a experiência efetuada para recolha de dados relativos a expressões faciais e valores referentes à presença de emoções no processo de audição de música.

3.2. Participantes

Os participantes nesta experiência constituem uma amostra de 26. Trata-se de uma amostra de conveniência. Participaram 6 elementos do sexo feminino com idades compreendidas entre os 19 e os 37 anos de idade e 20 elementos do sexo masculino com idades compreendidas entre os 19 e os 40 anos de idade. A média de idades dos participantes femininos é de 23 anos e dos participantes masculinos é de 28 anos. 20 participantes são estudantes na Faculdade de Engenharia da Universidade do Porto (FEUP), sendo 15 de diversas engenharias e 5 alunos do Mestrado Multimédia. Os restantes 6 participantes são externos à FEUP e não são estudantes.

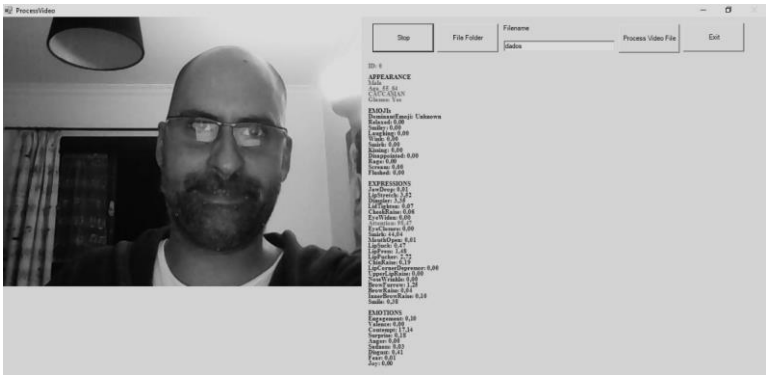
3.3. Materiais e equipamentos

Para encontrar uma relação entre o gosto musical manifestado emocionalmente sem recurso a nenhuma interface física em contacto com o ouvinte, torna-se imperativa a avaliação das expressões faciais. Como referido nos capítulos anteriores, é possível medir a emoção recorrendo a manifestações físicas da nossa cara como indicadores. Num sistema autónomo de recomendação essa avaliação terá que ser feita com recursos tecnológicos capazes de interpretar expressões faciais e registar valores indicadores de presença de emoções.

Há no mercado *softwares* comerciais que permitem realizar a avaliação de expressões faciais e emoções recorrendo a câmaras do tipo *webcam*, sensores de temperatura e batimentos cardíacos tais como; *Kairos Face Recognition, CrowdSight, Fraunhofer SHORE FaceDetect, Open Face Master, Noldus Face Reader, Emotient, Imotions*, mas são plataformas fechadas que não permitem o acesso aos dados. A marca *Affectiva* disponibiliza um *Software Development Kit* de nome *Affdex* que permite a integração dos dados registados com uma *webcam* em aplicações informáticas.

Uma aplicação para extrair os dados foi criada em *Microsoft Visual Studio* a partir do *SDK Affdex* (Figura 1) que permitiu iniciar a deteção das expressões, configurar o número de *frames* usados na deteção, e produzir um ficheiro de dados de extensão *CSV* para integração e análise com outros programas.

Figura 1.



Aplicação usada para extrair dados a partir do *SDK Affdex*

O número de *frames* permitido pelo *SDK* é de 30 *frames* por segundo sendo que foram registados cerca de 14 e 15 devido às capacidades da *webcam*. Alterando as condições de iluminação como teste, foi o máximo que se conseguiu registar.

O *software* usado para a reprodução dos excertos musicais, o *Foobar2000*, através de uma *interface* áudio ligado ao computador enviou o sinal a um sistema de colunas 2.1 (duas colunas e um *subwoofer*) amplificadas.

O número de excertos musicais foi escolhido em função do tempo praticável para este estudo totalizando cerca de 10 minutos de escuta. Esta distribuição resultou em 17 excertos musicais com aproximadamente 35 segundos cada um com exceção do primeiro do último com aproximadamente 40 segundos com a finalidade de servir de introdução e finalização à reprodução para a tornar menos abrupta. Cada excerto musical foi iniciado e terminado com um aumento e diminuição gradual do volume (*fade in* e *fade out*) para as transições não provocarem reações no comportamento cujos valores medidos fossem exagerados por diferenças muito grandes de volume do áudio. Este aspeto deve-se ao facto de as músicas não começarem no seu início e consequentemente não terem nenhuma parte introdutória. Para serem mais facilmente reconhecidas dentro do tempo estipulado optou-se por reproduzir os excertos num formato de verso seguido do refrão de cada tema musical. A exposição a entidades desconhecidas pode gerar atitudes neutras pelo menos momentaneamente (Woodside; & Chebat, 2001). O áudio para a reprodução foi editado com o *software Cubase 5* onde foi colocado cada excerto sem um critério específico de ordenação. Foi corrigido o volume de cada um para a reprodução ser o mais uniforme possível devido às variações inerentes à produção musical de cada registo.

Para promover uma reação emocional optou-se por escolher música referenciada como popular, tendo em conta os *sites* (acharts, 2018), (shazam, n.d.) e (billboard, n.d.), onde consta informação compilada referente a álbuns e singles mais ouvidos, para aumentar a probabilidade de ser conhecida dos participantes da experiência. Com base nessa informação indica-se na Tabela 1, a lista escolhida de músicas pela ordem da sua reprodução na recolha de dados nesta experiência.

Tabela 1

Ordem de reprodução das músicas selecionadas

Ordem	Artista/Banda	Nome da música
1	Mark Ronson	Uptown Funk (feat Bruno Mars)
2	Adele	Rolling In The Deep
3	Pharrell Williams	Happy
4	Joan Jett	I Love Rock'N'Roll
5	AcDc	Back In Black
6	Daft Punk	Get Lucky (feat Pharrell Williams)
7	Ed Sheeran	Castle On The Hill
8	Guns'N'Roses	Sweet Child O'Mine
9	Imagine Dragons	On Top Of The World
10	Megadeth	Symphony Of Destruction
11	Led Zeppelin	Whole Lotta Love
12	Kendrick Lamar	Loyalty (feat Rihanna)
13	Metallica	Enter Sandman
14	Nirvana	Smells Like Teen Spirit
15	Radiohead	Creep (radio edit)
16	Michael Jackson	Beat It
17	Queen	We Will Rock You

Para atribuição de um *rating* relacionado com o gosto pessoal de cada participante usou-se um questionário composto por três conjuntos de questões. O primeiro diz respeito ao conhecimento da música ouvida, com hipótese de resposta dicotómica (sim ou não); o segundo ao reconhecimento do artista ou banda, igualmente com hipótese de resposta dicotómica (sim ou não), e o terceiro continha uma escala tipo *Likert* de 7 valores, correspondendo o primeiro a não gostei da música e o sétimo a gostei muito da música.

3.4. Procedimentos

Um conjunto de 26 participantes, que autorizaram um pedido de consentimento informado referente à recolha de dados de vídeo, foram filmados durante a reprodução de excertos de 17 músicas num total de aproximadamente 10 minutos para cada um. Durante a filmagem as expressões faciais e as emoções presentes foram registadas para posterior análise. A cada participante foi pedido que se sentasse em frente a uma mesa com candeeiros para melhor iluminação do espaço como adjuvante ao funcionamento da captação vídeo. Em frente aos participantes em cima da mesa estava a *webcam* ligada a um computador portátil a correr o *Software Development Kit* e um *software* de captura de vídeo, *OBS Studio*, para ficar com o registo da experiência pois o *SDK* só regista os valores das expressões e emoções não gravando vídeo.

A sequência da implementação da recolha de dados foi a seguinte: o participante sentou-se em frente à mesa com o computador e a câmara, foi-lhe solicitado um minuto de respiração pausada, e que interpretasse aquele momento como uma situação calma de audição musical em ambiente caseiro. Ativou-se o *software* de captura de vídeo, iniciou-se a aplicação para reconhecimento de expressões faciais e seguidamente iniciou-se o programa com a sequência das músicas. A sequência das músicas era precedida por dois minutos de silêncio inicial para dar tempo de deixar o participante sozinho na sala antes do início do áudio. No final da sequência foi pedido ao participante que respondesse ao questionário.

Com recurso à aplicação criada, extraíram-se valores numéricos dos parâmetros de movimentos da face e valores que descrevem a presença mais ou menos acentuada das emoções. Os dados considerados neste estudo são relativos às emoções de *fear*, *anger*, *contempt*, *disgust*, *joy*, *sadness* e *surprise*. Foram extraídos valores correspondentes ao modelo *valence-arousal* cujas variáveis tomam respetivamente o nome de *valence* e *engagement*. Os valores numéricos das emoções variam de 0 a 100 excetuando a variável *valence* que varia de -100 a 100. O valor 0 significa ausência total da emoção e o valor 100 a sua presença completamente evidenciada. Os valores de *valence* significam o quão

agradável ou desagradável é o estado afetivo, isto é, o prazer sentido, e os valores de *engagement* que indicam o nível de estimulação sendo o máximo 100 e o mínimo 0. Foram considerados também os valores das expressões faciais descritas no *software: Attention, BrowFurrow, BrowRaise, CheekRaise, ChinRaise, Dimpler, EyeClosure, EyeWiden, InnerBrowRaise, JawDrop, LidTighten, LipCornerDepressor, LipPress, LipPucker, LipStretch, LipSuck, MouthOpen, NoseWrinkle, Smile, Smirk* e *UpperLipRaise*. As expressões faciais são medidas em valores de 0 a 100 sendo o 0 a ausência de expressão e o 100 a expressão muito presente.

Foi gerado um ficheiro *CSV* com os valores das variáveis descritas (referentes às emoções e expressões faciais), para cada um dos 26 participantes. Para separar os dados por música de cada participante recorreu-se à captura de ecrã para registar em vídeo o seu comportamento e perceber em que ponto temporal começa e acaba cada registo musical. O *software* usado, *OBS Studio,* produziu um ficheiro de vídeo onde se pode observar na aplicação criada com o *SDK* o comportamento dos participantes, os valores das variáveis em cada momento da gravação e no áudio, quando começa e acaba cada excerto musical. A partir deste vídeo procedeu-se à identificação dos valores das variáveis registadas no ficheiro *CSV* para cada música e esse conjunto foi separado em tabelas de *Excel*. Cada tabela referente a cada música ouvida por participante contém as variáveis indicadas acima por cada *frame* capturado em cada música bem como o registo do *frame* correspondente.

As respostas ao questionário foram dadas após a recolha de dados das expressões faciais sendo feita uma nova audição dos excertos para recordar a sua sequência e tentar identificar as músicas e artistas atribuindo-lhes uma pontuação referente ao gosto e prazer que cada participante experienciou durante a audição.

4. Resultados

4.1. Resumo

O objetivo da experiência desenvolvida para este trabalho é explorar a possibilidade de fazer algum tipo de previsão da avaliação dos ouvintes dada a cada música como contributo para inclusão em sistemas de recomendação de música automáticos. Vários cálculos foram considerados em busca de possíveis relações entre os valores medidos das varáveis de emoção, das variáveis do modelo *valence-arousal* e das expressões faciais retirados dos resultados do *Affdex SDK*.

4.2. Valores obtidos

Os valores obtidos em ficheiro *CSV* resultado da filmagem do comportamento dos participantes durante a audição dos excertos musicais, vêm discriminados por cada *frame* capturado e são referentes às variáveis descritas no capítulo anterior.

O processo de análise dos dados obtidos foi baseado na procura de correlações entre os valores medidos das diversas variáveis com os *ratings* atribuídos às músicas pelos participantes. Pretende-se desta forma determinar se alguma das variáveis medidas se encontra associada aos valores de *rating* recolhidos.

Numa perspetiva macroscópica de análise dos dados, foi calculada a média de todos os *ratings* atribuídos por todos os participantes a cada música. Por cada música das 17 usadas obtivemos a média das avaliações atribuídas pelos participantes.

Dois processos foram usados para correlacionar os valores da média dos *ratings* com as variáveis extraídas da experiência auditiva.

No primeiro processo, para cada música ouvida foram considerados os valores máximos de cada variável por participante. São os valores representativos das diversas emoções e expressões faciais mais fortes em cada música. Para cada variável foi calculada a média dos valores máximos por participante em cada música. Uma correlação de *Pearson* foi utilizada entre a média dos *ratings* e a média dos valores máximos de cada variável por pessoa por cada excerto musical.

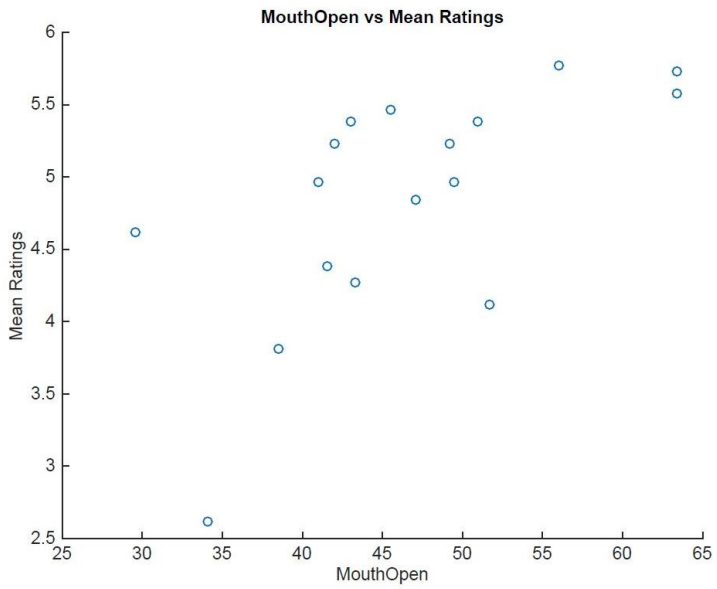
Os resultados indicaram correlações positivas e estatisticamente significativas para duas das variáveis medidas. Para a variável de expressão facial *MouthOpen* verificou-se uma correlação positiva $r = .63$, $p < .01$ e para a variável representativa da emoção *Disgust* uma correlação positiva $r = .61$, $p < .01$. Por este processo mais nenhuma correlação significativa foi encontrada.

No segundo processo substituíram-se os valores máximos de todas as variáveis pela mediana calculada a partir dos valores de cada variável durante o excerto musical correspondente. Para cada música calculou-se a mediana para cada participante relativa a cada variável calculando-se no fim o valor da média das medianas de todos os participantes para cada variável em cada música. A mediana de cada variável corresponde a um valor de tendência central, que não é tão influenciado pelos *outliers* quanto a média, de cada emoção sentida e de cada expressão facial durante um tema musical dando um panorama mais geral de cada emoção e expressão presentes na audição. Verificou-se uma correlação negativa da expressão facial *LipStretch*: $r = .85$, $p < .0001$.

No caso da avaliação referente à média calculada a partir dos valores máximos de cada variável, de todos os participantes para cada música, observámos duas correlações positivas estatisticamente significativas, sendo uma delas; uma emoção e outra uma expressão facial. A Figura 2 e a Figura 3 são gráficos do tipo *scatter plots*, representativos dos valores calculados e observa-se que há uma tendência de subida na média das classificações do grupo para cada um dos excertos musicais quanto maior é o valor da expressão *Mouth Open* em média do grupo de participantes, em cada música. Observa-se

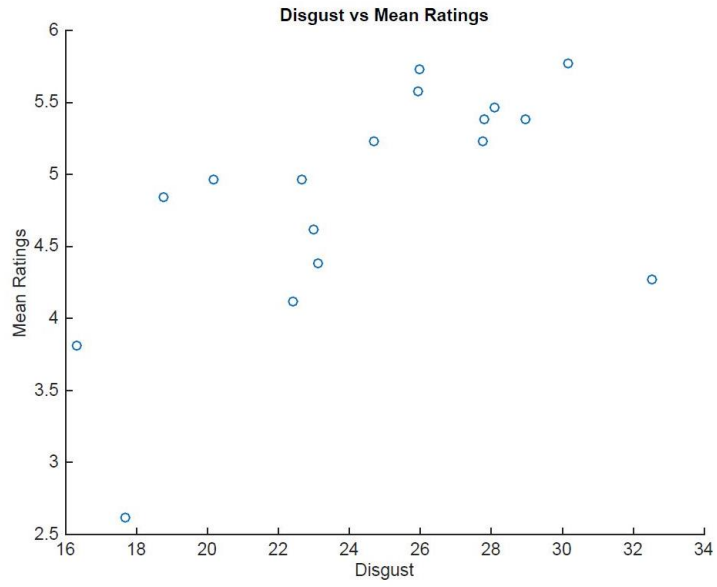
também uma tendência de subida da média de avaliação atribuída quanto maior é a média da emoção de *Disgust* do grupo de participantes numa música.

Figura 2.



Scatter plot da correlação entre a média *MouthOpen* e a média *Ratings*

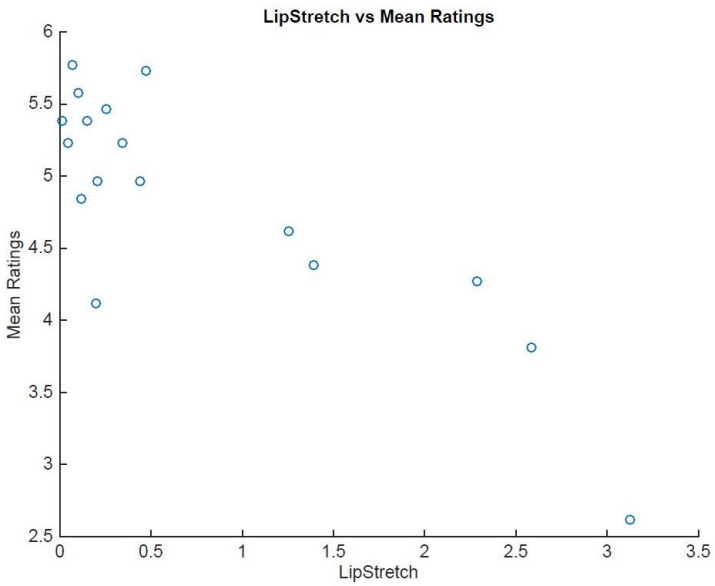
Figura 3.



Scatter plot da correlação entre a média *Disgust* e a média *Ratings*

A Figura 4 representa um gráfico do tipo *scatter plot* da correlação calculada de forma semelhante à descrita acima, mas a média calculada para todos os participantes por música foi obtida a partir da mediana dos valores medidos de cada variável durante o período de cada excerto musical. Desta forma descreve-se um valor central de cada variável não dando relevância às oscilações fortes dos valores medidos. Assim ignoram-se pequenos momentos musicais que possam provocar algum tipo de reação localizada em determinadas partes da música, tentando uniformizar as expressões e emoções medidas durante todo o trecho musical. Encontrou-se uma correlação forte e negativa de $r = .85$, $p < .0001$ da variável de expressão *LipStretch*.

Figura 4.



Scatter plot da correlação entre a média *LipStretch* e a média *Ratings*

Foram também calculadas correlações a nível individual. Para cada participante calculou-se a média de cada variável por cada música. Cada média de cada variável foi correlacionada com o rating atribuído a cada música. As Tabelas 2,3,4,5,6,7,8 e 9 mostram as correlações significativas obtidas.

Tabela 2

Correlações estatisticamente significativas entre variáveis de emoção e os participantes para $p < .05$

Participante	1	2	3	4	5	6	7	8	9	10	11	12	13
Fear	-	-	-	-	-	-	-	-	-	-	-	-	-
Anger	-	-	-	-	-	-	-	-	-	-	-	-	-
Contempt	-	-	-	-	-0,66	-	-	-	-	-	-	-	-
Disgust	0,52	-	-	-	-	-	-0,59	-	-	-	-	-	-
Joy	0,48	-	-	-	-	-	-	-	-	-	-	-	-
Sadness	-	0,72	-	-	-	-	-	-	-	-	-	-	0,54
Surprise	-	-	-	-	-	-	-	-	0,54	-	-	-	-
Engagement	-	-	-	-	-	-	-	-	-	-	-	-	-
Valence	0,57	-	-	-	-	-	-	-	-0,52	-	-	-	-

Tabela 3

Correlações estatisticamente significativas entre variáveis de emoção e os participantes para $p < .05$

Participante	14	15	16	17	18	19	20	21	22	23	24	25	26
Fear	-	-	-	-	-0,71	-	-	-	-	-	-	-	-
Anger	-	-	-	-	-	-	-	-	-	-	-	-	-
Contempt	0,6	-	-	-0,68	-0,59	-	-	-	0,76	-0,61	-	-	-
Disgust	-	-	-	-	-	-	-	-	0,59	-	0,50	-	-
Joy	-	-	-	-	-	-	-	-	-	-	-	-	-
Sadness	-	-	-	-	-	-	-	-	-	-	-	-	0,53
Surprise	-	-	-	-	-	-	-	-	-	-	-	0,50	-
Engagement	-	-	-	-	-	-	-	-	-	-	0,51	-	-
Valence	-	-	-	-	-	-	-	-	-	-	-	-	-

Tabela 4

Correlações estatisticamente significativas entre variáveis de expressões faciais e os participantes para p < .05

Participante	1	2	3	4	5	6	7	8	9	10	11	12	13
Attention	-	-	-	-	-	-	-	-	-	-	-	-	-
BrowFurrow	-	0,54	-	-	-	-	-	-	-	-	-	-	-
BrowRaise	-	-	-	-	-	-	-	-	-	-	-	-	-
CheekRaise	-	-	-	-	-	-	-	-	-	-	-	-	-
ChinRaise	-	-	-	-	-	-	-	-	-	-	0,52	-	-
Dimpler	-	-	-0,52	-	-	-	-	-	-	-	-	-	-
EyeClosure	-	-	-	-	-	-	-	-	0,51	-	-	-	-
EyeWiden	-	-	-	-	-	-	-	-	-	-	-	-	-
InnerBrow Raise	-	0,49	-	-	-	-	0,54	-	-	-	-	-	-
JawDrop	0,58	-	-	-	-	-	-	-	0,64	-	-	-	-
LidTighten	-	-	-	-	-	-	-	-	-	-	-	-	-
LipCorner Depressor	-	-	-	-	-	-	-	-	-	-	-	-	-
LipPress	-	-	-	-	-	-	-	-	-	-	-	-	-
LipPucker	-	-	-	-	-	-	-	-	-	-	-	-	-
LipStretch	-	-	-0,52	-	-	-	-	-	-0,61	-	0,53	-	-
LipSuck	-	-	-	-	-	-	-	-	-	-	-	-	0,58
MouthOpen	0,53	-	-0,48	-	-	-	-	-	0,76	-	-	-	0,50
Nose Wrinkle	-	-	-	-	-	0,54	-	-	-	-	-	-	-
Smile	0,49	-	-	-	-	-	-	-	-	-	-	-	-
Smirk	-	-	-0,52	-	-0,67	-	-	-	-	-	-	-	-
UpperLip Raise	0,49	-	-	-	-	-	-	-	-	-	-	-	-

Tabela 5

Correlações estatisticamente significativas entre variáveis de expressões faciais e os participantes para p < .05

Participante	14	15	16	17	18	19	20	21	22	23	24	25	26
Attention	-	-	-	-	-	-	-	-	-	-	-	-	-
BrowFurrow	-	-	-	-	-	-	-	-	-	-	-	0,53	-
BrowRaise	0,56	-	-	-	-	-	-	-	-	-	-	-	-
CheekRaise	-	-	-	-	-	-	-	-	-	-	0,57	-	-
ChinRaise	-	-	-	-	-	-	-	-	-	-	-	0,51	-
Dimpler	-	-	-	-	-	-	-	-	-	-	-	0,77	-
EyeClosure	-	-	-	-	-	-	-	-	-	-0,62	-	-0,49	-
EyeWiden	-	-	-	-	-	-	-	-	-	-	-	-	-
InnerBrow Raise	-	-	-	-	-	-	-	-	-	-	-	0,64	0,49
JawDrop	-	-	-	-	-	-	-	-	-	-0,55	-	-	-
LidTighten	-	-	-	-0,69	-	-	-	-	-	-	-	-	-
LipCorner Depressor	-	-	-	-	0,50	-0,54	-	-	-	-	-	-	-
LipPress	-	-	-	-	-	-	-	-	-	-	-	-	-
LipPucker	-	-	-	-	-	-	-	-	-	-	-	-	-
LipStretch	-	-	-	-	-	-	-	-	-	-	-	-	-
LipSuck	-	-	-	-	-	-	-	-	-	-	-	0,63	0,63
MouthOpen	0,48	-0,49	-	-0,49	-	-	-	-	0,71	-0,55	-	-	-
Nose Wrinkle	-	-	-	-	-	-	-	-	0,54	-	0,50	-	-
Smile	-	-	-	-	-	-	-	-	-	-	-	-	-
Smirk	0,53	-	-	-	-	-	-	-	0,52	-0,71	-	-	-
UpperLip Raise	-	-	-	-	-	-	-	-	-	-	-	-	-

Tabela 6

Correlações estatisticamente significativas entre variáveis de emoção e os participantes para p < .01

Participante	1	2	3	4	5	6	7	8	9	10	11	12	13
Fear	-	-	-	-	-	-	-	-	-	-	-	-	-
Anger	-	-	-	-	-	-	-	-	-	-	-	-	-
Contempt	-	-	-	-	-0,66	-	-	-	-	-	-	-	-
Disgust	-	-	-	-	-	-	-	-	-	-	-	-	-
Joy	-	-	-	-	-	-	-	-	-	-	-	-	-
Sadness	-	0,72	-	-	-	-	-	-	-	-	-	-	-
Surprise	-	-	-	-	-	-	-	-	-	-	-	-	-
Engagement	-	-	-	-	-	-	-	-	-	-	-	-	-
Valence	-	-	-	-	-	-	-	-	-	-	-	-	-

Tabela 7

Correlações significativas entre variáveis de emoção e os participantes para p < .01

Participante	14	15	16	17	18	19	20	21	22	23	24	25	26
Fear	-	-	-	-	-0,70	-	-	-	-	-	-	-	-
Anger	-	-	-	-	-	-	-	-	-	-	-	-	-
Contempt	-	-	-	-0,70	-	-	-	-	0,76	-0,60	-	-	-
Disgust	-	-	-	-	-	-	-	-	-	-	-	-	-
Joy	-	-	-	-	-	-	-	-	-	-	-	-	-
Sadness	-	-	-	-	-	-	-	-	-	-	-	-	-
Surprise	-	-	-	-	-	-	-	-	-	-	-	-	-
Engagement	-	-	-	-	-	-	-	-	-	-	-	-	-
Valence	-	-	-	-	-	-	-	-	-	-	-	-	-

Tabela 8

Correlações estatisticamente significativas entre variáveis de expressões faciais e os participantes para p < .01

Participante	1	2	3	4	5	6	7	8	9	10	11	12	13
Attention	-	-	-	-	-	-	-	-	-	-	-	-	-
BrowFurrow	-	-	-	-	-	-	-	-	-	-	-	-	-
BrowRaise	-	-	-	-	-	-	-	-	-	-	-	-	-
CheekRaise	-	-	-	-	-	-	-	-	-	-	-	-	-
ChinRaise	-	-	-	-	-	-	-	-	-	-	-	-	-
Dimpler	-	-	-	-	-	-	-	-	-	-	-	-	-
EyeClosure	-	-	-	-	-	-	-	-	-	-	-	-	-
EyeWiden	-	-	-	-	-	-	-	-	-	-	-	-	-
InnerBrow Raise	-	-	-	-	-	-	-	-	-	-	-	-	-
JawDrop	-	-	-	-	-	-	-	-	0,64	-	-	-	-
LidTighten	-	-	-	-	-	-	-	-	-	-	-	-	-
LipCorner Depressor	-	-	-	-	-	-	-	-	-	-	-	-	-
LipPress	-	-	-	-	-	-	-	-	-	-	-	-	-
LipPucker	-	-	-	-	-	-	-	-	-	-	-	-	-
LipStretch	-	-	-	-	-	-	-	-	-0,60	-	-	-	-
LipSuck	-	-	-	-	-	-	-	-	-	-	-	-	-
MouthOpen	-	-	-	-	-	-	-	-	0,76	-	-	-	-
Nose Wrinkle	-	-	-	-	-	-	-	-	-	-	-	-	-
Smile	-	-	-	-	-	-	-	-	-	-	-	-	-
Smirk	-	-	-	-	-0,70	-	-	-	-	-	-	-	-
UpperLip Raise	-	-	-	-	-	-	-	-	-	-	-	-	-

Tabela 9

Correlações estatisticamente significativas entre variáveis de expressões faciais e os participantes para p < .01

Participante	14	15	16	17	18	19	20	21	22	23	24	25	26
Attention	-	-	-	-	-	-	-	-	-	-	-	-	-
BrowFurrow	-	-	-	-	-	-	-	-	-	-	-	-	-
BrowRaise	-	-	-	-	-	-	-	-	-	-	-	-	-
CheekRaise	-	-	-	-	-	-	-	-	-	-	-	-	-
ChinRaise	-	-	-	-	-	-	-	-	-	-	-	-	-
Dimpler	-	-	-	-	-	-	-	-	-	-	-	0,77	-
EyeClosure	-	-	-	-	-	-	-	-	-	-0,60	-	-	-
EyeWiden	-	-	-	-	-	-	-	-	-	-	-	-	-
InnerBrow Raise	-	-	-	-	-	-	-	-	-	-	-	0,64	-
JawDrop	-	-	-	-	-	-	-	-	-	-	-	-	-
LidTighten	-	-	-	-0,7	-	-	-	-	-	-	-	-	-
LipCorner Depressor	-	-	-	-	-	-	-	-	-	-	-	-	-
LipPress	-	-	-	-	-	-	-	-	-	-	-	-	-
LipPucker	-	-	-	-	-	-	-	-	-	-	-	-	-
LipStretch	-	-	-	-	-	-	-	-	-	-	-	-	-
LipSuck	-	-	-	-	-	-	-	-	-	-	-	0,63	-
MouthOpen	-	-	-	-	-	-	-	-	0,71	-	-	-	-
Nose Wrinkle	-	-	-	-	-	-	-	-	-	-	-	-	-
Smile	-	-	-	-	-	-	-	-	-	-	-	-	-
Smirk	-	-	-	-	-	-	-	-	-	-0,70	-	-	-
UpperLip Raise	-	-	-	-	-	-	-	-	-	-	-	-	-

5. Discussão dos resultados

O estudo *afetivo* das reações humanas a estímulos musicais afirma-se hoje em dia como uma tendência (Kim et al., 2010) como descrito no capítulo 2.2. Os resultados indicados nos gráficos das Figuras 2 e 3 do capítulo 4.2 revelam que a média dos *ratings* de todos os utilizadores por cada excerto musical tem correlação com a expressão facial *MouthOpen* e a emoção de *Disgust* normalmente interpretada como uma emoção negativa. A expressão *MouthOpen* disponibilizada pelo *SDK Affdex* não está diretamente identificada no *Facial Action Coding System*. No entanto observando fotografias descritas como a emoção de *Disgust* os lábios podem estar abertos sem se tocarem (Ekman & Friesen, 1975) o que sugere que a emoção de *Disgust* é bom indicador global de avaliações mais elevadas da música tal como a expressão *MouthOpen*. A Figura 4 do capítulo 4.2 corresponde ao gráfico da correlação negativa entre as médias dos ratings de cada participante por cada excerto musical com as médias das variáveis. Ao contrário da avaliação anterior, não foram utilizados os valores máximos das variáveis, mas sim a mediana numa abordagem de uniformizar as expressões e emoções medidas durante cada música. Numa escala mais pequena cujos valores estão contidos num intervalo entre 0 e 3.5 (o *SDK* mede expressões entre os valores 0 e 100 sendo 0 ausência da expressão e 100 a sua total manifestação) um ligeiro movimento dos lábios (*LipStretch*) mostra-se como um bom indicador geral de não se gostar da música e consequentemente atribuir-lhe um *rating* mais baixo uma vez que a correlação foi alta e negativa. Este resultado demonstra que quando os participantes fazem essa expressão é um indicador que não gostam da música. O facto de o intervalo de valores registados ser tão pequeno significa que a presença dessa expressão é subtil, ou seja, o movimento não é muito acentuado. No entanto, verifica-se uma tendência para que quanto mais evidente é o movimento de *LipStretching* menor é o valor de avaliação da música.

Como não é expectável que os participantes atribuam *ratings* semelhantes a cada excerto musical nem que se registem emoções e expressões com as mesmas intensidades

(Ekman, 2004), foram avaliadas as correlações significativas de todas as variáveis para cada participante por cada música ouvida.

Os resultados das tabelas 2, 3, 4, 5, 6, 7, 8 e 9, do capítulo 4.2, não são conclusivos em relação às emoções como um bom indicador do gosto individual pela música. Não são também conclusivos em relação às expressões faciais. Os participantes 3, 4, 6, 8, 10, 11, 12, 15, 16, 19, 20 e 21 não manifestaram quaisquer emoções correlacionadas com a avaliação das músicas que fizeram. As emoções mais correlacionadas são a de *Contempt* e de *Disgust* verificando-se também alguma correlação para *Fear*, *Joy*, *Sadness* e *Surprise* mas num número menor de participantes. Observando a tabela 2 os participantes 14 e 22 apresentam correlações positivas e os participantes 5, 17, 18 e 23 apresentam correlações negativas referentes à emoção *Contempt*. Se para o primeiro conjunto de participantes referido as correlações positivas significam que *Contempt* indica gostarem da música, o mesmo não se pode afirmar em relação aos participantes com correlação negativa. Para a emoção *Disgust* os resultados são também antagónicos: os participantes 1, 22 e 24 apresentam correlações positivas e o participante 7 uma correlação negativa.

Os participantes 4, 8, 10, 12, 16, 20 e 21, não manifestaram qualquer correlação referente a expressões faciais. Observando a expressão *MouthOpen* das Tabelas 4, 5, 8 e 9, verifica-se que foi a que obteve mais correlações distribuídas pelos participantes. Concluímos que para alguns quanto mais evidente a expressão, menos gostaram da música (o *rating* atribuído foi mais baixo quanto mais presente a expressão resultando em correlação negativa) enquanto que para outros o resultado é oposto. Verifica-se que os movimentos relacionados com a boca e os lábios se mostram mais relevantes do que as expressões que contemplam movimentos relacionados com os olhos e sobrancelhas. Foram encontradas mais correlações para movimentos musculares da zona inferior da cara mais próximos da boca (*MouthOpen*, *LipStretch*, *LipSuck*, *LipCornerDepressor*, *JawDrop*, *Smirk*, *Dimpler*, *ChinRaise*, *Smile*, *UpperLipRaise*) do que da zona superior mais relacionados com os olhos e o nariz.

As assimetrias encontradas na distribuição das correlações por participante apelam a uma espécie de mapa de emoções e expressões faciais característico de cada indivíduo na forma como reage à música.

6. Conclusão e trabalho futuro

Quando se refere a recomendação *afetiva* de conteúdos audiovisuais faz-se um apelo ao estudo da emoção. Pretende-se estabelecer uma relação homem-máquina através de manifestações emocionais que devem ser bem interpretadas. As emoções apesar de poderem ser identificadas pela combinação da manifestação de expressões faciais compostas pelo movimento de vários músculos da cara, têm uma causa, um processo que as despoleta que não é mensurável.

No contexto da recomendação musical a avaliação da emoção revela-se insuficiente para descrever o prazer associado à escuta de música. O estado emocional de desprezo (*Disgust*) mostra-se como um bom indicador global do prazer associado à audição de algumas músicas sendo, porém, comummente interpretado como uma emoção negativa.

As expressões faciais quando avaliadas isoladamente umas das outras, não as relacionando diretamente com estados emotivos específicos, mostram-se como indicadores interessantes da reação dos ouvintes à música. Quando consideradas como *AUs* (*Action Units*) individuais *MouthOpen* revela-se como um bom indicador médio para o grupo de participantes de gosto pela música e a *AU LipStretch* como um sinal subtil de desagrado.

Se as correlações encontradas parecem bons indicadores médios para o grupo de participantes principalmente no respeitante às expressões faciais onde a boca e lábios têm um papel preponderante, as correlações para cada indivíduo participante mostram que a emoção não é garante de uma boa interpretação. A nível individual duas das emoções mais relevantes que obtiveram correlação, *Disgust* e *Contempt*, são suscetíveis de serem confundidas com outras. A emoção *Disgust* pode confundir-se com *Anger* ou com *Contempt* em função da combinação das *Action Units* presentes e a sua intensidade. *Fear* e *Sadness* que também têm alguma correlação são consideradas emoções negativas e podem ser identificadas só com movimentos relacionados com os olhos e sobrancelhas não sendo obrigatória a presença de movimentos dos lábios e boca. Estes antagonismos

não estabelecem uma correlação clara com as atitudes perante a música. A manifestação emocional pode depender de fatores culturais e pode revelar-se com diferentes intensidades de indivíduo para indivíduo o que não favorece uma universalização da sua interpretação. A causa que provoca um estado emocional enquanto se ouve música não pode ser medida, mas um conjunto de expressões faciais isoladas desenham um "mapa" das características de cada indivíduo. Foram encontradas mais correlações relativas a expressões faciais do que a estados emotivos (sete participantes não têm expressões faciais correlacionadas e doze participantes não têm emoções correlacionadas), mas em combinações diferentes para cada participante. A expressão de *MouthOpen* surge como o melhor indicador de apreciação de música, mas não exclusivamente. Verifica-se que movimentos da boca e lábios e zonas musculares mais próximas são mais relevantes do que movimentos dos músculos próximos aos olhos e sobrancelhas. No entanto existe uma idiossincrasia na resposta facial à audição de música.

Não se conseguindo prever qual o conjunto de expressões pertencentes a um só indivíduo que favorecem uma atitude de avaliação em relação à música, consegue-se fazer um apelo a sistemas do tipo *Machine Learning* que interpretem os conjuntos de expressões faciais de cada um aprendendo, como indicam as correlações efetuadas neste trabalho, a partir de que momento um ouvinte se desinteressa por uma música.

As referências atuais apontam para um caminho de estudo *afetivo* em relação aos conteúdos audiovisuais. O termo *afetivo* ou *affective* tem sido associado a estados emotivos que são genericamente estudados a partir da combinação das várias expressões faciais designadas como *Action Units*. Este trabalho considerando conteúdos musicais, evidencia uma tendência para o estudo das expressões de forma individualizada, não as relacionando diretamente com os estados emotivos como um processo a explorar sendo os movimentos relacionados com a boca e lábios bons indicadores de preferência musical que potenciam a interpretação da vontade de mudar de música por parte do ouvinte.

Como trabalho futuro seria interessante fazer a avaliação de um só participante em mais do que uma sessão para verificar o seu estado emocional em diferentes dias face às mesmas músicas. Verificar se as correlações das expressões faciais individuais se mantêm confrontando-as com possíveis correlações dos estados emocionais. Alterar a

sequência musical para tentar perceber se tem influência na manifestação emocional e se dependendo do dia em que é feita a avaliação a atitude face à música se mantém. Reforçar os resultados em relação a movimentos da boca e lábios como bons indicadores de preferência musical. Recomendar música baseado nos resultados.

7. Referências

Baltrusaitis, T. (2014). Automatic facial expression analysis. *University of Cambridge*, 220. https://doi.org/10.4018/978-1-59140-562-7.ch010

Bogdanov, D., & Herrera, P. (2011). How Much Metadata Do We Need in Music Recommendation? A Subjective Evaluation Using Preference Sets. *12th International Society for Music Information Retrieval Conference (ISMIR'11), Proc.*, (Ismir), 97–102.

Cano, P., Koppenberger, M., & Wack, N. (2005). Content-based music audio recommendation. *Proceedings of the 13th Annual ACM International Conference on Multimedia - MULTIMEDIA '05*, (January 2005), 211. https://doi.org/10.1145/1101149.1101181

Chordia, P., Godfrey, M., & Rae, A. (2008). Extending content-based recommendation: the case of Indian classical music. *International Conference on Music Information Retrieval*, 571–576. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:EXTENDING+CONTENT-BASED+RECOMMENDATION+:+THE+CASE+OF+INDIAN+CLASSICAL+MUSIC#0

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. *The Handbook of Emotion Elicitation and Assessment*, 203–221. https://doi.org/10.1007/978-3-540-72348-6_1

Dalton, M. K., Ferraro, E. J., Galuardi, M., Robinson, M. L., Stauffer, A. M., & Walls, M. T. (2016). ON THE INCORPORATION OF PSYCHOLOGICALLY-DRIVEN "MUSIC" PREFERENCE MODELS FOR MUSIC RECOMMENDATION. *University of Maryland, College Park*.

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, *111*(15), E1454–E1462. https://doi.org/10.1073/pnas.1322355111

Ekman, P. (2004). *Emotions revealed*. *Student BMJ* (Vol. 12). https://doi.org/10.1002/bsl

Ekman, P. (2016). What Scientists Who Study Emotion Agree About. *Perspectives on Psychological Science*, *11*(1), 31–34. https://doi.org/10.1177/1745691615596992

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues*. *Journal of Personality*. https://doi.org/10.1163/1574-9347_bnp_e804940

Elias Pampalk Tim, Tim Pohle, G. W. (2005). Dynamic Playlist Generation Based On Skipping Behavior. *Proc. of the 6th ISMIR Conference*, (January), 634--637. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.59.7880

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, *22*(1), 5–53. https://doi.org/10.1145/963770.963772

Kanoje, S., Girase, S., & Mukhopadhyay, D. (2014). User Profiling Trends, Techniques and Applications. *International Journal of Advance Foundation and Research in Computer*, *1*(11), 2348–4853.

Keltner, D., Ekman, P., Gonzaga, G. C., & Beer, J. (2003). Facial expression of emotion. *Handbook of Affective Sciences*. https://doi.org/10.1177/1754073910361979

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., … Turnbull, D. (2010). Music Emotion Recognition : a State of the Art Review. *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, (Ismir), 255–266. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.231.7740&rep=rep1&type=pdf%5Cnhttp://ismir2010.ismir.net/proceedings/ismir2010-45.pdf

Pasi, S. (2015). Music Mood Annotation Using Semantic Computing and Machine Learning. *Faculty of Humanities of the University of Jyväskylä*. Retrieved from https://jyx.jyu.fi/dspace/bitstream/handle/123456789/45096/978-951-39-6074-2.pdf?sequence=3

Salomon, A., & Logan, B. (2001). A Content-Based Music Similarity Function. *Cambridge Research Laboratory 2001/02*, (June).

Singh, G., & Boparai, R. S. (2016). A Novel Hybrid Music Recommendation System using K-Means Clustering and PLSA. *Indian Journal of Science and Technology*, *9*(28). https://doi.org/10.17485/ijst/2016/v9i28/95592

Soleymani, M., Asghari-Esfeden, S., Fu, Y., & Pantic, M. (2016). Analysis of EEG Signals and Facial Expressions for Continuous Emotion Detection. *IEEE Transactions on Affective Computing*, *7*(1), 17–28. https://doi.org/10.1109/TAFFC.2015.2436926

Song, Y., Dixon, S., & Pearce, M. (2012). A survey of music recommendation systems and future perspectives. *9th International Symposium on Computer Music Modeling and Retrieval*, (June), 19–22. Retrieved from http://cmmr2012.eecs.qmul.ac.uk/sites/cmmr2012.eecs.qmul.ac.uk/files/pdf/papers/cmmr2012_submission_42.pdf

Wang, J., de Vries, A. P., & Reinders, M. J. T. (2006). Unifying user-based and item-based collaborative filtering approaches by similarity fusion. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '06*, 501. https://doi.org/10.1145/1148170.1148257

Woodside, A. G., & Chebat, J. C. (2001). *Updating Heider's balance theory in consumer behavior: A Jewish couple buys a German car and additional buying-consuming transformation stories*. *Psychology and Marketing* (Vol. 18). https://doi.org/10.1002/mar.1017

Hard rock albums (n.d.). Retirado de https://www.billboard.com/charts/hard-rock-albums https://acharts.co/portugal_albums_top_30/2018/09. (n.d.).

https://www.shazam.com/pt/charts/hall-of-fame. (n.d.).

https://www.billboard.com/charts/greatest-billboard-200-albums. (n.d.).

https://www.billboard.com/charts/artist-100. (n.d.).

https://www.billboard.com/charts/greatest-hot-100-singles. (n.d.).

https://www.billboard.com/charts/rock-albums. (n.d.).

8. Anexo - Questionário

Questionário sobre experiência musical

(Carvalho, J., Davies, M., & Moreira, L. – 2018)

Nome: _____

Idade: _____ Género: Masc. ☐ Fem. ☐

Email: _____

1 - Conhece a música que ouviu? (assinale com X na coluna correspondente)

	Sim	Não
Música 1		
Música 2		
Música 3		
Música 4		
Música 5		
Música 6		
Música 7		
Música 8		
Música 9		
Música 10		
Música 11		
Música 12		
Música 13		
Música 14		
Música 15		
Música 16		
Música 17		

2 - Consegue identificar o artista/banda correspondente a cada música?

	Sim	Não
Música 1		
Música 2		
Música 3		
Música 4		
Música 5		
Música 6		
Música 7		
Música 8		
Música 9		
Música 10		
Música 11		
Música 12		
Música 13		
Música 14		
Música 15		
Música 16		
Música 17		

FEUP Mestrado Multimédia – Tecnologias 2018 João Pedro Carvalho – jpdiascarvalho@gmail.com

3 - Atribua uma classificação a cada música. 1 significa que não gostou da música e 7 que gostou muito. Coloque um X na coluna correspondente.

	1	2	3	4	5	6	7
Música 1							
Música 2							
Música 3							
Música 4							
Música 5							
Música 6							
Música 7							
Música 8							
Música 9							
Música 10							
Música 11							
Música 12							
Música 13							
Música 14							
Música 15							
Música 16							
Música 17							