

**UNIVERSIDADE FEDERAL DE SÃO CARLOS**

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**M-MOTION: UM TOCADOR DE MÚSICA  
DIGITAL PARA PLATAFORMA MÓVEL QUE  
CONSIDERA EMOÇÕES DOS USUÁRIOS**

**WILLIAN GARCIAS DE ASSUNÇÃO**

**ORIENTADORA: PROF. DR. VÂNIA PAULA DE ALMEIDA NERIS**

São Carlos – SP

Outubro, 2018

**UNIVERSIDADE FEDERAL DE SÃO CARLOS**

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**M-MOTION: UM TOCADOR DE MÚSICA  
DIGITAL PARA PLATAFORMA MÓVEL QUE  
CONSIDERA EMOÇÕES DOS USUÁRIOS**

**WILLIAN GARCIAS DE ASSUNÇÃO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação, área de concentração: Interação Humano-Computador.

Orientadora: Prof. Dr. Vânia Paula de Almeida Neris

São Carlos – SP

Outubro, 2018



# UNIVERSIDADE FEDERAL DE SÃO CARLOS

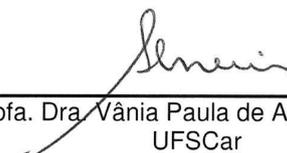
Centro de Ciências Exatas e de Tecnologia  
Programa de Pós-Graduação em Ciência da Computação

---

## Folha de Aprovação

---

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Willian Garcias de Assunção, realizada em 24/10/2018:



---

Profa. Dra. Vânia Paula de Almeida Neris  
UFSCar



---

Prof. Dr. Cesar Augusto Camillo Teixeira  
UFSCar

---

Prof. Dr. Marcelo Soares Pimenta  
UFRGS

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Marcelo Soares Pimenta e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ao) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.



---

Profa. Dra. Vânia Paula de Almeida Neris

*Este trabalho é dedicado aos meus pais, Iron e Juscelina, à minha irmã Larissa e aos meus avós,  
Cassimira e Orestes.*

# AGRADECIMENTOS

---

---

Primeiramente eu gostaria de agradecer a Deus, pela força, coragem, luz e paciência em tantos momentos difíceis.

A toda minha família, pelo grande incentivo e por todo carinho. Em especial ao meu pai, minha mãe, minha irmã e meus avós.

Além do mais, gostaria de ressaltar que este trabalho não poderia ter sido realizado se não fosse o apoio, direto ou indireto, de muitas pessoas que estiveram ao meu lado ou passaram pelo meu caminho ao longo desses anos. Não poderei citar aqui o nome de todos, mas em alguns em especial gostaria de agradecer.

Agradeço aos professores da Universidade Federal de Goiás, que contribuíram para minha formação, bem como não mediram esforços para que eu pudesse ingressar o mais rápido possível na UFSCar. Em especial ao professor Marcos Wagner de Souza Ribeiro pelo incentivo em entrar no programa de mestrado no segundo semestre de 2016. Ao amigo Renan Vinícius Aranha, que desde a graduação me apoia, incentiva e dá suporte para continuar nessa caminhada. E também a minha amiga Andressa Moraes, que em todos os dias buscou me alertar para a escrita da qualificação de mestrado, mesmo em dias que eu estava com baixo estado de ânimo. Não poderia esquecer também de todas as pessoas que participaram dos experimentos dessa pesquisa, pois, acredito que sem essas pessoas este trabalho não poderia ter seguido adiante, portanto, deixo meu imenso agradecimento pela colaboração do meu trabalho bem como do avanço da ciência.

Agradeço ao PPGCC da UFSCar e à todo o corpo docente pela oportunidade que me deram de desenvolver este trabalho e pela valiosa formação. Agradeço também pela concessão de uma bolsa CAPES que me permitiu dedicar grande parte do tempo à realização desta pesquisa.

Por fim, agradeço à professora e orientadora Vânia Paula de Almeida Neris, pela sua confiança, incentivo e transparência em todos os momentos. Além de suas ideias e sugestões fornecidas durante as atividades deste mestrado.

*"Paciência e perseverança tem o efeito  
mágico de fazer as dificuldades desapa-  
recerem e os obstáculos sumirem"*  
(John Quincy Adams)

## RESUMO

Nos últimos anos houve um avanço no desenvolvimento de tecnologias que digitalizaram a experiência de escuta da música. Com isso, um grande acervo de música tornou-se disponível na internet, além do surgimento de novas técnicas automatizadas de seleção de música. Os sistemas atuais de reprodução de música recuperam as músicas por metadados como: título, autor, banda e gênero musical. Uma vez que a música pode transmitir informações relacionadas à emoção, estudos mostram que a música é um meio eficaz de indução emocional e pode mudar o comportamento emocional do usuário. Os métodos atuais de recomendação e reprodução de música dependendo da emoção requer uma interação manual do usuário e muitas das vezes são sugeridas a partir de um conjunto de músicas já pré-classificadas, por esse motivo, isso se torna um problema para o usuário, visto que muitas das vezes o usuário tem uma tarefa intensiva em selecionar uma música de acordo com a emoção desejada diante de um grande conjunto de músicas diversificadas e de estilos variados. Sendo assim, este trabalho tem como objetivo propor o m-Motion, uma ferramenta de reprodução musical para smartphones que busca ajudar o usuário a alcançar um estado emocional desejado partindo do seu estado emocional atual. A emoção atual e desejada do usuário são obtidas pela expressão de sentimentos subjetivos, informados por meio do discurso da voz (texto) e de um elemento de interface de usuário inspirado no espaço semântico de Scherer. As emoções das músicas do usuário são mapeadas para um espaço dimensional de excitação e valência, que a princípio são previstas pelo algoritmo de classificação Support Vector Machine (SVR). Diante disso, este trabalho adotou o espaço semântico de Scherer para classificar as emoções das músicas e dos usuários em excitação e valência. Sendo assim, o algoritmo do m-Motion faz o uso de duas fórmulas matemáticas, sendo a distância euclidiana e a equação da reta entre a emoção atual e desejada. Com isso o algoritmo pode retornar um conjunto de músicas entre a emoção atual e desejada do usuário. Um experimento foi realizado com três grupos distintos de usuários que tinha como objetivo avaliar a emoção alcançada pelos usuários após a reprodução de conjunto de músicas sugeridas. Os dois primeiros grupos tinham 20 usuários cada experimento, portanto, de cada usuário foi coletado o estado emocional atual e desejado, além das expressões faciais durante a reprodução das músicas. O primeiro grupo recebeu sugestões de músicas retornadas pelo algoritmo proposto, já o segundo grupo as músicas eram sugeridas manualmente pelo próprio usuário. O terceiro grupo continha 8 usuários e o experimento foi realizado fora do ambiente controlado. Os usuários deste grupo utilizaram a aplicação m-Motion durante 5 dias e receberam sugestões de músicas com base ao estado emocional desejado. Os resultados sugerem que, ao ouvir as músicas selecionadas pela aplicação proposta, os usuários podem se aproximar do estado emocional desejado, bem como da recomendação manual realizada pelo próprio usuário.

**Palavras-chave:** Reprodutor de música baseado em emoções, recomendação de música, emoção musical, reconhecimento de emoção musical, recuperação de música por emoção.

# ABSTRACT

In recent years there has been a breakthrough in the development of technologies that have digitized the listening experience of music. With this, a large collection of music became available on the internet, in addition to the emergence of new automated techniques of music selection. Current music playback systems retrieve music through metadata such as title, author, band, and music genre. Since music can convey information related to emotion, studies show that music is an effective means of emotional induction and can change the emotional behavior of the user. The current methods of music recommendation and playback depending on the emotion requires a manual user interaction and are often suggested from a set of already pre-sorted songs, so this becomes a problem for the user, since most of the time the user has an intensive task in selecting a song according to the desired emotion before a large set of diverse songs and varied styles. Thus, this work aims to propose m-Motion, a musical reproduction tool for smartphones that seeks to help the user achieve a desired emotional state from their current emotional state. The current and desired emotion of the user is obtained by the expression of subjective feelings, informed through the speech of the voice (text) and of an element of user interface inspired in the semantic space of Scherer. The emotion of the user's songs are mapped to a dimensional space of excitation and valence, which are predicted by the classification algorithm Support Vector Machine (SVR). Therefore, this work adopted the semantic space of Scherer to classify the emotions of the songs and the users in terms of excitement and valence. Thus, the m-Motion algorithm makes use of two mathematical formulas, being the Euclidean distance and the equation of the line between the current and desired emotion. With this, the algorithm is able to return a set of songs between the current and desired emotion of the user. An experiment was conducted with three distinct groups of users that aimed to assess the emotion achieved by users after playing a set of suggested songs. The first two groups had 20 users each experiment and therefore each user was collected the current and desired emotional state, as well as the facial expressions during music playback. The first group received suggestions of songs returned by the proposed algorithm, while the second group the songs were manually suggested by the user. The third group contained 8 users and the experiment was carried out outside the controlled environment. Users of this group used the m-Motion application for 5 days and received music suggestions based on the desired emotional state. The results suggest that, when listening to the songs selected by the proposed application, the users can approach the desired emotional state, as well as the manual recommendation made by the user.

**Keywords:** Music player based on emotions, recommendation of music, musical emotion, recognition of musical emotion, recovery of music by emotion.

## LISTA DE SIGLAS

---

---

API	Interface de Programação de Aplicações
BPM	Batidas por Minuto
CNNs	Convolutional neural networks
DWCH	Daubechies WaveletCoefficient Histograms
FMA	Free Music Archive
HRI	Human-Robot Interaction
IEEE	Institute of Electrical and Eletronics Engineers
IHC	Interção Humano-Computador
LIFES	Laboratório de Interação Flexível e Sustentável
MER	Music Emotion Recognition
MFCCs	Cepstrum de Mel-Frequência
MIDI	Musical Instrument Digital
OMCS	Oracle Management Cloud
OSC	Octave-based Spectral Contrast
P2P	Peer-to-Peer
PCM	Pulse-Code Modulation
PDA	Personal Digital Assistant
RAV	Ressonância-Excitação-Valência
RGB	Red Green and Blue
RMSE	Root Mean Square Error
SAM	Self Assessment Manikin

SGD	Descenramento Estocástico de Gradiente
SSDs	Statistical Spectrum Descriptors
SVR	Support Vector Regression
XPOD	A Human Activity and Emotion Aware Mobile Music Player
WAV	WAVEform audio format

## LISTA DE FIGURAS

---

---

Figura 1 – Espaço Emocional Semântico de Scherer . . . . .	24
Figura 2 – Processo de Reconhecimento de Emoção em Música . . . . .	26
Figura 3 – Processo de extração de recursos de dados da música. . . . .	27
Figura 4 – Cartão Self-Assessment Manikin (SAM) utilizado nos experimentos. . . . .	31
Figura 5 – XPOD interface do usuário. . . . .	34
Figura 6 – XPOD Formato da rede neural. . . . .	34
Figura 7 – Interface HeartPlayer. . . . .	35
Figura 8 – HeartPlayer em execução. . . . .	35
Figura 9 – LyQ: Arquitetura. . . . .	37
Figura 10 – LyQ: Diagrama de fluxo para recuperação de recursos de emoção. . . . .	38
Figura 11 – LyQ: Interface. . . . .	38
Figura 12 – Diagrama de blocos da arquitetura do sistema emotion-aware music player. . . . .	39
Figura 13 – Captura da tela do player de música. . . . .	40
Figura 14 – Arquitetura do Sistema. . . . .	41
Figura 15 – Captura da tela de treinamento do sistema. . . . .	42
Figura 16 – Interface EmoPlayer. . . . .	44
Figura 17 – Módulo de recomendação. . . . .	45
Figura 18 – Arquitetura m-Motion. . . . .	50
Figura 19 – Análise das músicas do usuário por meio do Servidor do LIFeS. . . . .	51
Figura 20 – Funcionamento do servidor do LIFeS. . . . .	52
Figura 21 – Dados de entrada dos usuários capturados pelo m-Motion. . . . .	53
Figura 22 – Recuperação da emoção do usuário por meio do texto da voz. . . . .	54
Figura 23 – Recuperação dos metadados das músicas do usuário. . . . .	54
Figura 24 – Funcionamento do aplicativo m-Motion . . . . .	55
Figura 25 – Comportamento da função <i>dPointLine</i> . . . . .	56
Figura 26 – Comportamento da função <i>limit</i> . . . . .	58
Figura 27 – Etapas para o algoritmo para selecionar músicas. . . . .	59
Figura 28 – Etapas para o algoritmo para selecionar músicas. . . . .	60
Figura 29 – Interface visual para captura do texto da voz do m-Motion. . . . .	61
Figura 30 – Componente visual Scherer. Adaptado do modelo do espaço semântico de Scherer para captura da emoção atual e desejada via dispositivo móvel. . . . .	62

Figura 31 – Interface da lista de reprodução do m-Motion. . . . .	63
Figura 32 – Componente visual Scherer. Adaptado do modelo do espaço semântico de Scherer para captura da emoção alcançada do usuário via dispositivo móvel. . . . .	64
Figura 33 – Componente visual SAM. Adaptado da técnica self-assessment manikin (SAM) para captura da emoção alcançada do usuário via dispositivo móvel. . . . .	65
Figura 34 – Interface web do espaço semântico. . . . .	69
Figura 35 – Interface web para captura de emoção por meio das expressões faciais. . . . .	72
Figura 36 – Total acumulativo da excitação ao longo da reprodução das listas sugeridas para cada participante. . . . .	72
Figura 37 – Total acumulativo da valência ao longo da reprodução das listas sugeridas para cada participante. . . . .	73
Figura 38 – Comparação dos níveis de excitação alcançada pelos participantes antes e após ouvir as músicas sugeridas pelo algoritmo. . . . .	74
Figura 39 – Comparação dos níveis de valência alcançada pelos participantes antes e após ouvir as músicas sugeridas pelo algoritmo. . . . .	74
Figura 40 – Total acumulativo da excitação ao longo da reprodução das listas sugeridas para cada participante. . . . .	76
Figura 41 – Total acumulativo da valência ao longo da reprodução das listas sugeridas para cada participante. . . . .	76
Figura 42 – Comparação dos níveis de excitação alcançada pelos participantes antes e após ouvir as músicas sugeridas pelo algoritmo. . . . .	77
Figura 43 – Comparação dos níveis de valência alcançada pelos participantes antes e após ouvir as músicas sugeridas pelo algoritmo. . . . .	78
Figura 44 – Participante 1 - Níveis de Valência e Excitação. . . . .	80
Figura 45 – Participante 2 - Níveis de Valência e Excitação. . . . .	81
Figura 46 – Participante 3 - Níveis de Valência e Excitação. . . . .	82
Figura 47 – Participante 4 - Níveis de Valência e Excitação. . . . .	83
Figura 48 – Participante 5 - Níveis de Valência e Excitação. . . . .	85
Figura 49 – Participante 6 - Níveis de Valência e Excitação. . . . .	86
Figura 50 – Participante 7 - Níveis de Valência e Excitação. . . . .	87
Figura 51 – Participante 8 - Níveis de Valência e Excitação. . . . .	88

## LISTA DE TABELAS

---

---

Tabela 1 – Características descritivas comumente usadas no MER . . . . .	28
Tabela 2 – Comparativo de trabalhos relacionados. . . . .	47
Tabela 3 – Número total de recursos de curto prazo implementados pela biblioteca pyAudioAnalysis. . . . .	52
Tabela 4 – Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer. . . . .	75
Tabela 5 – Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer. . . . .	79
Tabela 6 – Participante 1 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer. . . . .	81
Tabela 7 – Participante 2 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer. . . . .	82
Tabela 8 – Participante 3 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer. . . . .	83
Tabela 9 – Participante 4 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer. . . . .	84
Tabela 10 – Participante 5 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer. . . . .	85
Tabela 11 – Participante 6 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer. . . . .	86
Tabela 12 – Participante 7 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer. . . . .	87
Tabela 13 – Participante 8 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer. . . . .	88
Tabela 14 – Meios de publicação . . . . .	91

# SUMÁRIO

---

---

<b>CAPÍTULO 1–INTRODUÇÃO</b> . . . . .	<b>16</b>
1.1 Contexto . . . . .	16
1.2 Motivação, Problemática e Objetivos . . . . .	17
1.3 Síntese da Metodologia . . . . .	18
1.4 Organização do Trabalho . . . . .	19
<b>CAPÍTULO 2–REFERENCIAL TEÓRICO</b> . . . . .	<b>21</b>
2.1 Considerações Iniciais . . . . .	21
2.2 Emoções . . . . .	21
2.2.1 Modelos para a Descrição das Emoções . . . . .	22
2.3 Música e Emoção . . . . .	23
2.3.1 Reconhecimento de Emoção Musical . . . . .	25
2.3.2 Extração de Recursos da Música . . . . .	26
2.3.3 Classificação das Características da Música . . . . .	28
2.3.4 Métodos de Reconhecimento de Emoção Musical . . . . .	30
2.4 Avaliação da Experiência do Usuário . . . . .	30
<b>CAPÍTULO 3–TRABALHOS RELACIONADOS</b> . . . . .	<b>32</b>
3.1 Considerações Iniciais . . . . .	32
3.2 XPOD – A Human Activity and Emotion Aware Mobile Music Player . . . . .	33
3.3 HeartPlayer:A Smart Music Player Involving Emotion Recognition, Expression and Recommendation . . . . .	34
3.4 LyQ - An Emotion-aware Music Player . . . . .	36
3.5 Embedded Design of an Emotion-Aware Music Player . . . . .	38
3.6 Facial Expression Based Music Player . . . . .	40
3.7 Emotion-based Music Recommendation Using Audio Features and User Playlist . . . . .	42
3.8 Emotion Based Mood Enhancing Music Recommendation . . . . .	43
3.9 Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation . . . . .	44
3.10 Comparação dos Trabalhos Relacionados . . . . .	45
<b>CAPÍTULO 4–M-MOTION</b> . . . . .	<b>49</b>
4.1 Considerações Iniciais . . . . .	49
4.2 Arquitetura . . . . .	49
4.3 Algoritmo . . . . .	55
4.4 Reprodutor mobile . . . . .	60

4.4.1	Interface de usuário para recuperação da emoção atual e desejada do usuário . . . . .	60
4.4.2	Interface de usuário da lista de reprodução . . . . .	62
4.4.3	Interface de usuário de avaliação da emoção alcançada . . . . .	63
<b>CAPÍTULO 5–AVALIAÇÃO . . . . .</b>		<b>66</b>
5.1	Considerações Iniciais . . . . .	66
5.2	Planejamento dos Experimentos . . . . .	66
5.2.1	Critérios de Inclusão de Usuários nos Experimentos . . . . .	67
5.2.2	Riscos em Participar dos Experimentos . . . . .	68
5.2.3	Benefícios em Participar dos Experimentos . . . . .	68
5.3	Recomendação pelo Algoritmo e Usuário . . . . .	68
5.3.1	Emoção Atual e Desejada do Participante . . . . .	69
5.3.2	Reconhecimento Emocional da Música . . . . .	69
5.3.2.1	Lista de Reprodução dos Participantes . . . . .	70
5.3.2.2	Extração de características . . . . .	70
5.3.2.3	Dataset . . . . .	70
5.3.3	Avaliação da emoção alcançada . . . . .	70
5.3.4	Resultados da Recomendação pelo Algoritmo . . . . .	71
5.3.5	Resultados da Recomendação Manual do Participante . . . . .	76
5.4	Recomendação pelo Reprodutor m-Motion . . . . .	78
5.4.1	Emoção Atual e Desejada do Usuário . . . . .	78
5.4.2	Reconhecimento Emocional da Música . . . . .	79
5.4.3	Avaliação da Emoção Alcançada . . . . .	79
5.4.4	Resultados da Recomendação pela Aplicação m-Motion . . . . .	80
<b>CAPÍTULO 6–CONCLUSÕES . . . . .</b>		<b>89</b>
6.1	Análise Crítica . . . . .	89
6.2	Publicações . . . . .	91
6.3	Trabalhos Futuros . . . . .	91
<b>REFERÊNCIAS . . . . .</b>		<b>92</b>
<b>A–TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO . . . . .</b>		<b>98</b>
<b>B–AUTORIZAÇÃO DE CAPTAÇÃO E EXIBIÇÃO DE IMAGEM, SOM E NOME . . . . .</b>		<b>100</b>
<b>C–ESPAÇO EMOCIONAL SEMÂNTICO DE SCHERER . . . . .</b>		<b>102</b>

**D-SELF-ASSESSMENT MANIKIN (SAM) . . . . . 103**

# Capítulo 1

## INTRODUÇÃO

---

---

### 1.1 Contexto

Segundo [Britannica \(2016\)](#), a música é uma arte que preocupa em combinar sons vocais ou instrumentais de forma organizada e de acordo com os padrões culturais de ritmo, melodia e harmonia. Qualquer evento intencionalmente produzido ou organizado para ser ouvido, e que tenha algum recurso musical básico pode se admitir como um som de música ([MCKEOWN-GREEN, 2014](#)).

A música é algo que está intrinsecamente interligada com a vida das pessoas e a maioria das pessoas desfruta da música em seu cotidiano. Ouvimos música na estrada, trabalhando, limpando a casa ou enquanto desfrutamos de um livro. Além disso, preferimos diferentes tipos de música para diferentes tipos de ocasiões ([JANSSEN et al., 2012](#)).

Ao longo dos últimos 20 anos, novas tecnologias digitalizaram nossa experiência de escuta musical. No mundo de hoje, temos acesso a grandes quantidades de música em todos os lugares e o tempo todo. No momento, há cada vez mais músicas em computadores pessoais, em bibliotecas de música e na Internet ([JANSSEN et al., 2012](#)).

Com o avanço da internet e o desenvolvimento crescente de smartphones <sup>1</sup>, as pessoas passaram a preferir ouvir músicas pela internet, pois podem escolher músicas de acordo com sua preferência ([WANG, 2007](#)). Com isso, houve um aumento na quantidade de consumo de música pela internet.

Ter acesso a grandes quantidades de música digital leva a demanda de técnicas automatizadas para a seleção de música. Na indústria da música, as demandas de recuperação e recomendação de música atraem muitos pesquisadores para explorar as características acústicas para a classificação de gêneros musicais ([LIN et al., 2016](#)).

Embora as informações tradicionais, como o artista, o álbum ou o título de um trabalho

---

<sup>1</sup> Smartphone é um telefone celular que tem um processador e computa. O termo em português faz referência a um celular "inteligente"

musical, continuem importantes, essas tags<sup>2</sup> possuem aplicabilidade limitada em muitas aplicações relacionadas à música (LU et al., 2006), por exemplo, os sistemas P2P (Peer-to-Peer) que não detêm controle de vocabulário na representação da música. A recuperação é feita pelo próprio usuário, normalmente pelo nome da música, dos autores ou álbum. Essa situação traz uma redundância, imprecisão e homogeneização nos processos de recuperação da informação da música, visto que as músicas apresentam conteúdos culturais e linguísticos para sua recuperação (SANTINI; SOUZA, 2007).

## 1.2 Motivação, Problemática e Objetivos

Uma das formas de otimizar o processo de seleção de música é aproveitar o poder emocional da música. Se um reprodutor de música conseguisse reconhecer a emoção das músicas, poderia gerar automaticamente listas de reprodução que despertassem, relaxassem ou tornassem o ouvinte mais alegre (JANSSEN et al., 2012). Por exemplo, quando um indivíduo volta para casa do trabalho, ele pode querer ouvir música relaxante e leve; enquanto ele está no ginásio, ele pode querer escolher uma música emocionante com uma batida forte e ritmo rápido. Em suma, muitas pessoas selecionam músicas considerando seu estado emocional.

Desta forma, com a capacidade de recuperar a emoção da música o tocador de música poderia se concentrar nas recomendações de músicas por emoção em que algumas atividades específicas exigem. Além disso, esta tecnologia poderá dar a possibilidade de sintonizar a emoção do usuário, ajustando a música selecionada ao seu estado emocional atual ou desejado.

De acordo com um relatório recente, a prevalência da maioria das atividades de lazer, como assistir televisão ou filmes, ou ler livros, foi superada pela música ouvida (RENTFROW; GOSLING, 2003). Quase todas as peças de música são criadas para transmitir sentimentos, por exemplo, compositores criam música para ressoar com seus ouvintes e os artistas usam os sons musicais para provocar as respostas emocionais do público (YANG et al., 2017).

Nos últimos anos, houve um aumento significativo de pesquisas sobre o tema de emoções na área de IHC, pois, a emoção pode auxiliar na tomada de decisões das pessoas. O estudo de Zentner et al. (2008) mostrou que a música é um meio eficaz de indução emocional e pode mudar o comportamento emocional do usuário. Além do mais, os pesquisadores também mostraram que, na vida cotidiana, a música é predominantemente usada para regulação do estado emocional.

No entanto, os métodos atuais de recomendação e reprodução de música dependendo da emoção do usuário requerem uma interação manual, ou seja, o usuário procura manualmente uma música de acordo com sua emoção ou recebe sugestões um conjunto de músicas já pré-classificadas, mas que podem não pertencer ao seu estilo musical preferido, conforme os trabalhos de Dornbush et al. (2005), Cervantes e Song (2013) e Kamble e Kulkarni (2016). Essa tarefa de buscar uma música com base a emoção desejada é intensiva para o usuário, pois, muitas das vezes

<sup>2</sup> Tag: em inglês significa etiqueta, rótulo.

enfrenta-se uma indecisão por se deparar com um grande conjunto de músicas diversificadas e de estilos variados.

A literatura também apresenta trabalhos que recomendam músicas com base a emoção atual usuário, por exemplo, os trabalhos de [Fan et al. \(2011\)](#), [Deng e Leung \(2012\)](#), [Iyer et al. \(2017\)](#) e [Gilda et al. \(2017\)](#). Porém, estes trabalhos a seleção de músicas que é executada geram uma lista de reprodução, de modo que selecionam músicas que refletem a emoção atual, em vez de melhorar a emoção do usuário. Então, se o usuário estiver triste, será retornado um conjunto de músicas com uma emoção triste que podem degradar ainda mais o seu humor e levá-lo à depressão.

Portanto, o presente trabalho de dissertação tem como objetivo propor uma aplicação móvel para reprodução de música que ajuda o usuário a alcançar um estado emocional desejado. As emoções são obtidas pela expressão de sentimentos subjetivos, informados por meio do discurso da voz (texto) e de um elemento de interface de usuário inspirado no espaço semântico de Scherer. A base do reprodutor consiste em um algoritmo que verifica o estado emocional do usuário e, então recomenda uma sequência de músicas do próprio usuário que partem do seu estado atual até o estado emocional desejado.

Diante disso, tendo essa pesquisa um propósito de buscar e analisar a influência da música com a emoção do usuário, e observando a utilização da aplicação proposta, a hipótese levantada por este trabalho é: se ouvir as músicas selecionadas pelo reprodutor proposto, então o usuário alcançará um estado emocional próximo ao desejado.

### 1.3 Síntese da Metodologia

Nesta dissertação de mestrado são propostos: a) a investigação de métodos para se inferir o estado emocional do usuário e para a classificação de músicas considerando o estado emocional que evocam e b) a adaptação da lista de músicas do usuário, considerando o estado emocional desejado do ouvinte e o estado emocional atual. Trata-se, portanto, de um trabalho de pesquisa exploratória<sup>3</sup>, que envolve o desenvolvimento de uma solução de software, e que visa a criação de uma nova infraestrutura (algoritmos, fluxo de dados e responsabilidades).

Este trabalho partiu de uma revisão bibliográfica que buscou elencar trabalhos que tinham como objetivo a automatização da interação entre um reprodutor mobile<sup>4</sup> e um usuário, de forma que essa automatização pudesse ser feita através das emoções das músicas do usuário. Além disso, verificaram-se trabalhos que também propuseram levar o usuário para um estado emocional desejado partindo do seu estado emocional atual utilizando alguma ferramenta.

<sup>3</sup> Pesquisa exploratória: É um tipo de pesquisa que tem como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a construir hipóteses, envolvem: (a) levantamento bibliográfico; (b) entrevistas com pessoas que tiveram experiências com o problema pesquisado; e (c) análise de exemplos que estimulem a compreensão.

<sup>4</sup> Reprodutor mobile: reprodutor de música para celulares.

A partir deste estudo, foi possível construir a aplicação para smartphones m-Motion que captura a emoção do usuário e adapta sua lista de música, de tal forma que os efeitos das músicas selecionadas pelo usuário possam ajudá-lo a atingir um estado emocional desejado.

A infraestrutura do sistema m-Motion possui sua implementação final como um serviço de streaming<sup>5</sup> de música convencional. No entanto, apenas os metadados das músicas do usuário estão contidos no servidor, bem como toda etapa de processamento e recuperação da emoção da música. Com isso, a infraestrutura da aplicação m-Motion foi concebida pensando no respeito aos direitos autorais e considerando que o usuário tenha uma lista de músicas prediletas.

Durante o desenvolvimento deste trabalho foi realizado um experimento com 3 grupos distintos de participantes, sendo dois grupos com 20 participantes e um grupo com 8 participantes. Os experimentos consistiam em analisar as emoções alcançadas pelos participantes após as recomendações de músicas feitas pelo algoritmo proposto, a aplicação m-Motion e o próprio participante. Além de verificar se o usuário aproximava-se do estado emocional desejado foi possível comparar se as recomendações realizadas pelo algoritmo e a aplicação m-Motion puderam aproximar-se da recomendação manual do próprio participante.

A avaliação da emoção do usuário foi realizada de três maneiras, primeiro utilizando a técnica de FACS (Facial Action Coding System), em seguida o SAM (Self Assessment Manikin) (BRADLEY; LANG, 1994) e por fim o espaço emocional semântico de Scherer (SCHERER, 2005) como artefato de registro.

## 1.4 Organização do Trabalho

O restante deste documento está dividido da seguinte forma:

- No Capítulo 2 são apresentados os conceitos de emoções, o espaço semântico de Scherer, uma revisão da literatura sobre a relação existente entre emoção e música, bem como as formas de reconhecimento de emoção musical.
- No Capítulo 3 são descritos oito trabalhos selecionados na literatura como sendo os mais próximos e relevantes para este trabalho, bem como é apresentada uma comparação entre eles.
- No Capítulo 4 é apresentada uma visão geral da aplicação m-Motion, sendo detalhada sua arquitetura, e descrito os comportamentos do algoritmo de recomendação e do reprodutor mobile.
- No Capítulo 5 é discutida a avaliação do m-Motion com experimentos realizados com a recomendação pelo algoritmo, seleção manual do usuário e da própria aplicação m-Motion em um ambiente real.

<sup>5</sup> Streaming: É uma tecnologia de transmissão instantânea de dados de áudio ou vídeo através de redes.

- Por fim, no Capítulo 6 conclui-se este documento apresentando uma análise crítica dos resultados.

# Capítulo 2

## REFERENCIAL TEÓRICO

---

---

### 2.1 Considerações Iniciais

Neste capítulo, é apresentada uma síntese do estudo da literatura sobre fatores emocionais na área de Interação Humano-Computador (IHC). Este capítulo está organizado da seguinte forma: a Seção 2.2 apresenta o conceito de Emoções e o modelo escolhido para classificá-las: Espaço semântico de Scherer, por fim a Seção 2.3 apresenta os conceitos de emoções no âmbito de sinais de áudio, mais precisamente para as músicas, bem como seus métodos de recuperação de emoção.

### 2.2 Emoções

Ao longo das últimas décadas, tornou-se crescente a pretensão de pesquisas no âmbito de aspecto emocional, com o intuito de cada vez mais buscar encontrar respostas que expliquem a diversidade da experiência emocional humana. Para [Scherer \(2001\)](#), a emoção é o processo que está relacionado às mudanças sincronizadas nas relações de todos ou a maioria dos diferentes componentes do organismo.

Para muitos psicólogos a emoção é tida como uma componente fundamental dos seres humanos, pois, segundo [Agarwal e Meyer \(2009\)](#) é impossível agir ou pensar sem envolver emoções. Além do mais, as emoções têm um papel essencial para os indivíduos avaliarem situações como sendo boas ou ruins ([NORMAN, 2005](#)).

A emoção é bastante estudada dentro da área de psicologia, no entanto, nem mesmo entre os pesquisadores da área é possível encontrar um consenso sobre a definição exata de emoção. Os primeiros estudos sobre os aspectos emocionais foram realizados por Darwin em 1872, Sócrates no século VI a.C., Platão nos anos de 427 a 347 a.C. e Tzu nos anos de 470 a 399 a.C ([SCHERER, 2001](#)).

Segundo [Scherer \(2009\)](#), uma emoção pode ser definida como um mecanismo de adaptação cultural e psicobiológica que permite que cada indivíduo reaja de forma flexível e dinâmica

às contingências ambientais. Ainda de acordo com [Scherer \(2005\)](#), a emoção envolve diferentes componentes, nomeadamente uma reação observável, uma excitação fisiológica, uma interpretação cognitiva, uma expressão motora e uma experiência subjetiva.

[Norman \(2004\)](#) defende que o afeto pode ser definido como um termo geral para o sistema de julgamento, seja consciente ou subconsciente, e a emoção pode ser entendida como a experiência consciente do afeto. Norman e Scherer ainda destacam a importância de se estudar e considerar as emoções do usuário nos processos de avaliação e decisões de design. Na área de IHC, a emoção vem sendo tratada como uma componente essencial da experiência com o usuário, pois, está diretamente relacionada ao processo de avaliação e nas decisões de design.

Com isso, encontrar uma definição para emoção não é uma tarefa trivial, por se tratar de um assunto complexo que gera discussão até mesmo entre os pesquisadores da área de psicologia. O que pode se concluir é que não existe definição certa, errada ou única para cada um dos termos ([GONCALVES, 2016](#)).

### 2.2.1 Modelos para a Descrição das Emoções

Buscando melhor compreender as emoções, diversos modelos que descrevem como as pessoas vivenciam suas emoções foram propostos e são encontrados na literatura. Alguns modelos descrevem as emoções utilizando a abordagem cognitiva ([ORTONY et al., 1988](#)) e outros consideram um espaço multidimensional como domínios de valência e excitação ([RUSSELL, 1980](#)). Há também modelos que utilizam uma abordagem de multicamadas (em inglês *process-level*) ([NORMAN, 2004](#)), ou por uma abordagem de avaliação ([DESMET et al., 2003](#)). É possível também encontrar modelos que são designados para contextos específicos de emoções, por exemplo, aquela provocada por produtos.

De maneira geral, os modelos que descrevem as emoções são divididos em duas perspectivas: uma perspectiva dimensional e uma perspectiva discreta ([LOCKNER et al., 2014](#)). Na perspectiva discreta as emoções são vistas como uma soma de categorias que podem ser separadas para obter subcategorias mais refinadas e menores. Diversos modelos foram propostos, por exemplo, [Plutchik e Kellerman \(1980\)](#) considera oito emoções (alegria/tristeza, confiança/nojo, medo/raiva, surpresa/antecipação) com base na sua capacidade de desencadear um comportamento de luta ou fuga. Os modelos discretos são bastante populares, isso porque eles são facilmente ligados à "psicologia popular" ([LOCKNER et al., 2014](#)).

No entanto, os modelos discretos apresentam algumas desvantagens. Vários estudos mostram que uma emoção pode ser difícil de categorizar, por exemplo, no trabalho de [Barrett e Wager \(2006\)](#). No trabalho, os autores relatam que um modelo discreto limita o número potencial de emoções, impedindo qualquer identificação mais profunda, precisa e induzindo vieses.

Por essas razões, outros modelos de emoções coexistem, baseados em uma perspectiva dimensional. Entre os pesquisadores, o número de dimensões varia. No entanto, duas dimen-

sões emergem da maioria os modelos dimensionais: valência e excitação (RUSSELL, 1980; BARRETT; WAGER, 2006; SCHERER, 2005). Valência corresponde a uma escala de prazer/desprazer, enquanto a excitação corresponde a uma escala de baixa excitação/alta excitação. Essas escalas definem um espaço circunflexo onde é possível localizar qualquer termo emocional de "psicologia popular".

Para Scherer (2005), muitas vezes não está claro se um julgamento de valência diz respeito à avaliação da natureza do objeto ou evento de estímulo, ou melhor, do sentimento induzido por ele. Da mesma forma, as classificações de excitação ou ativação podem se referir à ativação percebida em uma situação (ou imagem) ou à sensação proprioceptiva de excitação fisiológica induzida pelo evento de estímulo. Partindo disso, Scherer (2005) propôs um novo modelo de espaço emocional semântico que estende o modelo de (RUSSELL, 1980) adicionando dois novos domínios importantes para a diferenciar a emoção, os domínios de sentimento de controle e facilidade de conclusão do objetivo. Diante disso, este trabalho adotou o modelo de Scherer para descrição das emoções, que além da inclusão de dois novos domínios, o espaço semântico proposto por Scherer incluí oitenta termos emocionais comuns da "psicologia popular", veja a Figura 1.

A estrutura do espaço emocional semântico é dividida em quatro domínios:

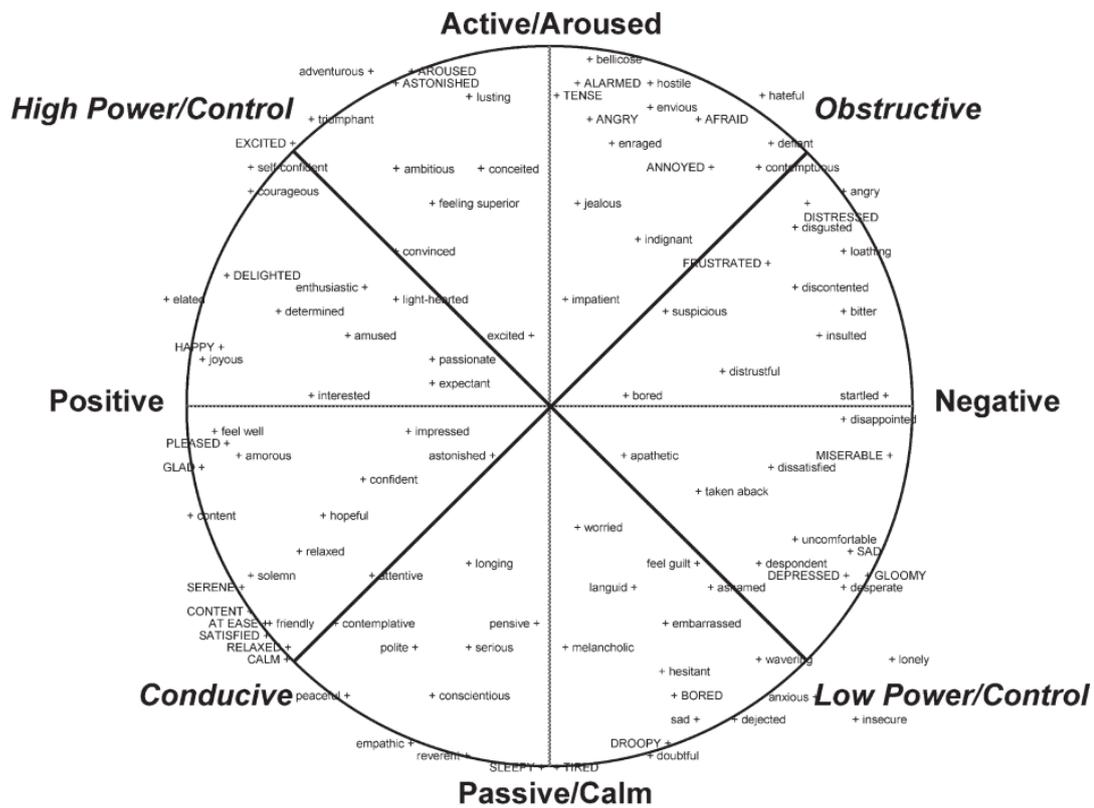
- Excitação (Arousal): domínio fortemente relacionado ao impacto de excitação e motivação do usuário;
- Sentimento de controle (Sense of control): reflete o sentimento de controle ou equilíbrio emocional do usuário sobre determinado objeto, situação ou circunstância;
- Valência (Valence): representa o grau de satisfação, prazer e desprazer;
- Facilidade de conclusão do objetivo (Ease in achieving a goal): está relacionado a facilidade e a dificuldade que o usuário possui em alcançar e concluir uma meta.

Outro aspecto que influencia na escolha do modelo dimensional de Scherer é que o modelo multidimensional é focado na experiência subjetiva do usuário. Como sentimentos subjetivos são restritos à experiência consciente das emoções, a identificação das respostas emocionais geralmente é obtida por instrumentos de autorrelato como questionários, o que pode torná-lo um modelo de avaliação das emoções de baixo custo e de rápida aplicação (SCHERER, 2000).

## 2.3 Música e Emoção

O som pode ser definido como a vibração do ar, ou seja, variações na pressão do ar que percebemos pelos nossos ouvidos. Por exemplo, ao tocar uma campinha uma energia

**Figura 1 – Espaço Emocional Semântico de Scherer**



Fonte: (SCHERER, 2005).

cinética é produzida, assim temos em um curto prazo uma "deformação" dessa campanha. Essa "deformação" produz vibrações que formam mudanças na pressão do ar. Com essa mudança na pressão do ar, ondas são emitidas aos ouvidos humano e recebidas pelo tímpano de maneira que possam ser escutadas na forma de sons (MILETTO et al., 2004). De acordo com Miletto et al. (2004), os sons quando são convertidos digitalmente são definidos como áudio, ou seja, o som é um fenômeno físico ou mecânico, enquanto o áudio é a sua representação eletrônica.

A seguir são apresentados alguns elementos básicos que compõem um áudio (MILETTO et al., 2004) (BENNETT, 1987):

- **Altura tonal:** é por meio da altura tonal que podemos diferenciar um som agudo de outro som grave. Por exemplo, ao tocar um piano, percebe-se que os sons mais agudos são localizados mais a direita enquanto os mais graves localizados a esquerda. Essa "altura" de um som que popularmente é chamada, se define como altura tonal.
- **Volume:** a altura de uma onda é chamada de amplitude e quanto maior a amplitude mais forte será o som. Portanto, o volume do som é determinado pela amplitude (altura da onda).
- **Intensidade:** é a força do som, também chamada de sonoridade. Com a intensidade é possível diferenciar se o som é fraco ou forte e está relacionada à energia de vibração da

fonte que emite as ondas sonoras. Quanto maior a energia que a onda transporta, maior é a intensidade que o ouvido pode perceber. A intensidade é um elemento semelhante ao Volume.

- **Timbre:** dois instrumentos musicais não produzem o mesmo som ao serem tocados com a mesma altura tonal e o mesmo volume. Isso porque existe um elemento a mais que os diferenciam e é chamado de timbre. Cada instrumento possui um timbre único, assim como cada pessoa também possui um timbre único de voz.
- **Harmonia:** é a combinação dos sons ouvidos simultaneamente, ou seja, o agrupamento agradável de sons.
- **Melodia:** é uma sequência de sons em intervalos irregulares, portanto, é dita como a sequência de notas musicais, que de forma organizada proporciona um sentido musical para quem está ouvindo.
- **Ritmo:** o ritmo é quem dita o tempo musical e o estilo da música. É a definição de quanto tempo cada parte da melodia irá continuar.

### 2.3.1 Reconhecimento de Emoção Musical

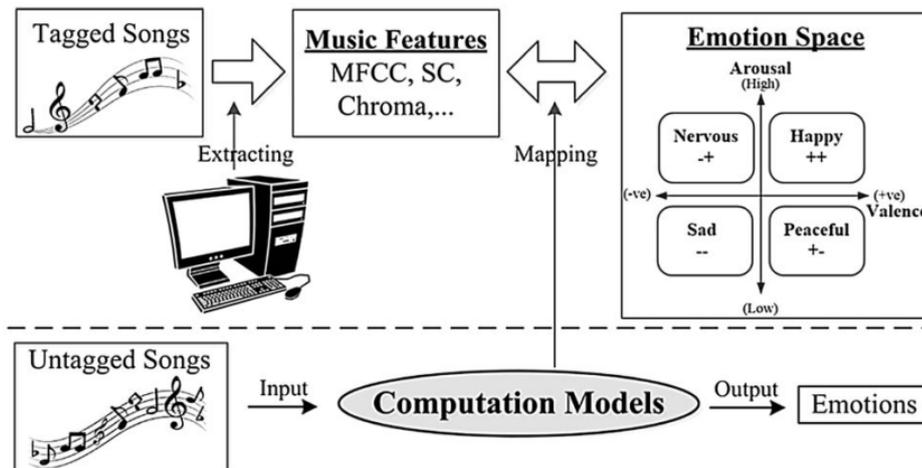
A emoção é a essência da música e essa informação pode ser utilizada para recuperar e recomendar uma música. Quase todos os componentes das músicas são criados para transmitir sentimentos. [Leonard \(1956\)](#) foi um musicólogo pioneiro cujo livro de 1956, "Emotion and Meaning in Music" buscou apresentar a existência de ligações entre respostas emocionais e padrões musicais, demonstrando que o sentido da música é intra-musical e que este é equivalente à expectativa.

No decorrer dos anos, técnicas para organizar e recuperar músicas usando uma abordagem baseada em emoção vêm se tornando viáveis e o núcleo de tal técnica é o reconhecimento automático da emoção. Atualmente, os métodos para usar as características físicas, auditivas e semânticas dos sinais de música para obter a informação emocional da música tornaram-se uma parte importante da pesquisa em aplicações de música digital ([DOBASHI et al., 2015](#)) ([YANG et al., 2017](#)).

A área de pesquisa que é responsável por investigar os modelos de computação para detectar a emoção nas músicas é conhecida como reconhecimento de emoção musical (*MER - Music Emotion Recognition*). Na [Figura 2](#) pode ser visto o processo de funcionamento do MER. O processo utiliza computadores para extrair e analisar características da música, para que assim possa realizar um mapeamento entre os recursos de áudio obtidos da música com o espaço emocional definido e assim reconhecer a emoção que a música expressa, de modo que os bancos de dados possam ser organizados e gerenciados com base na emoção ([KIM et al., 2010](#)).

O MER compreende três etapas ([BARTOSZEWSKI et al., 2008](#)) ([YANG et al., 2007](#)):

Figura 2 – Processo de Reconhecimento de Emoção em Música



Fonte: (YANG et al., 2017).

1. **Reconhecimento do domínio:** é o passo inicial que consiste em definir qual o formato do registro da música e o modelo de descrição de emoção para utilizar.
2. **Extração de informações:** nessa etapa é realizado o processo de extração de informações ou características sonoras do sinal de áudio e é essencial para obter a emoção da música.
3. **Classificação:** dados extraídos no passo anterior criam as entradas para os algoritmos de classificação. O conjunto de dados extraídos juntamente com o conjunto de treinamento desempenham um papel fundamental para a obtenção do resultado da classificação.

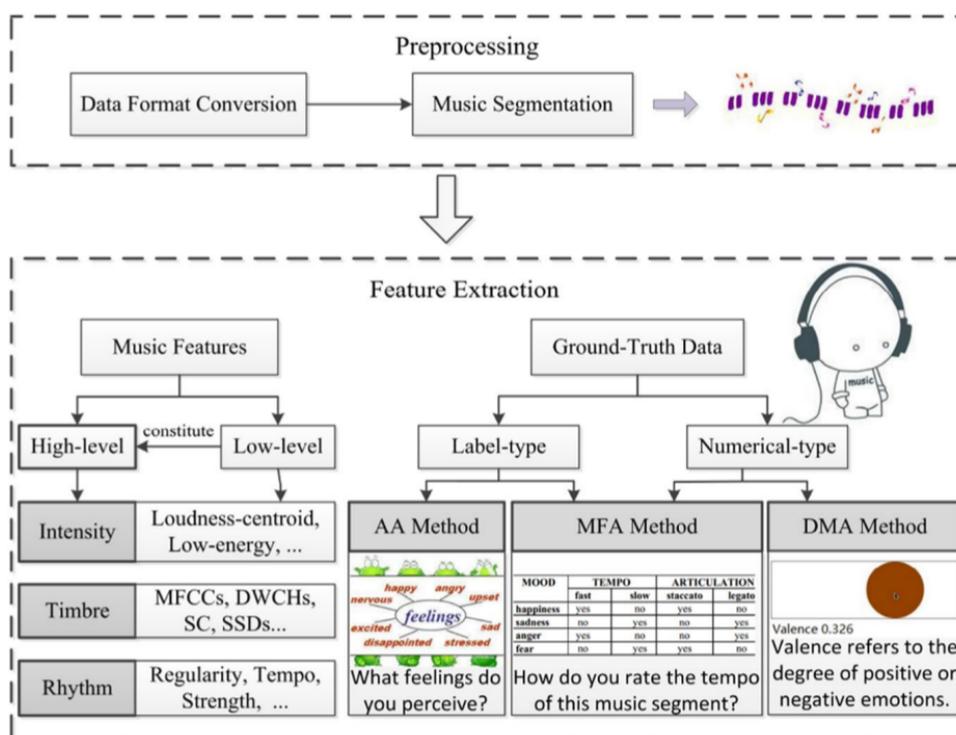
### 2.3.2 Extração de Recursos da Música

O processo de extração de recursos de dados busca reduzir as informações das músicas, de modo que os descritores possam descrevê-las completamente. Os recursos de dados utilizados no MER podem ser divididos em duas categorias: o primeiro é chamado de Recursos da Música, que é representado pelos recursos acústicos e as letras da música; o segundo é chamado de dados *ground-truth*, que são rótulos de emoção atribuídos pelos ouvintes que refletem verdadeiramente a emoção percebida pelos seres humanos (SOLEYMANI et al., 2013).

Como é mostrado na Figura 3, antes do processo de extração de recursos, os arquivos de músicas devem ser convertidos para um formato fácil de serem processados e em seguida os dois tipos de recursos de dados podem ser extraídos utilizando um software de extração específico ou métodos de anotações. O pré-processamento da música compreende de duas etapas: conversão do formato de dados e segmentação da música.

- Primeiro, o objetivo da conversão do formato de dados é converter os arquivos de áudio que armazenam a música em um formato que tenha as informações necessárias. Por

**Figura 3 – Processo de extração de recursos de dados da música.**



Fonte: (YANG et al., 2017).

exemplo, nos trabalhos de Yang et al. (2007) e Zhang et al. (2017) os arquivos de áudio foram convertidos para um formato uniforme (22,050 Hz, 16 bits e PCM (Pulse-Code Modulation) WAV (WAVEform audio format) de canal mono) e Bartoszewski converteu arquivos MIDI (Musical Instrument Digital) para um formato uniforme que contenha o tom, a duração e a velocidade da voz (BARTOSZEWSKI et al., 2008).

- Em segundo lugar, por mais que a emoção da música possa ser prevista pela estimativa geral de toda a música, a precisão dos resultados é baixa. Isso porque o conteúdo emocional da música varia no decorrer da música e a avaliação direta da música na sua totalidade pode levar a um erro. Portanto, para obter resultados mais precisos, primeiro é necessário dividir a música em pequenos segmentos e, em seguida, detectar a emoção separadamente (BARTOSZEWSKI et al., 2008).

Diferentes métodos de segmentação da música diferem entre si, sendo que cada um possui suas próprias vantagens e desvantagens, além do mais o tamanho de cada segmento também se difere. MacDorman (2007) em seu trabalho relatou que o comprimento dos segmentos devem ser o mais curto possível, pois, quanto menor o segmento mais homogêneo será a emoção e isso tornará os resultados da avaliação dos indivíduos mais consistentes.

No entanto, um segmento muito curto não é suficiente para permitir avaliar que seu conteúdo emocional tenha uma boa precisão e de acordo com Wang et al. (2012), o comprimento do segmento para a música popular geralmente é de 25 a 30s, o que corresponde ao comprimento

típico da parte do refrão dele e para músicas clássicas [Xiao et al. \(2008\)](#) argumenta em seu trabalho que o comprimento típico do segmento obteve melhores resultados para segmentos com tamanho entre 8 e 16s.

### 2.3.3 Classificação das Características da Música

Alguns estudos mostram que diferentes estados emocionais geralmente são associados a diferentes características musicais ([GABRIELSSON; LINDSTRÖM, 2001](#)) ([LU et al., 2006](#)) ([THAYER; MCNALLY, 1992](#)). Por exemplo, o estado emocional "contentamento" está relacionado a um tempo musical baixo, com baixa intensidade e um timbre suave, enquanto o estado emocional "exuberância" está relacionado com um ritmo mais rápido, com alta intensidade e de timbre alegre. Além do mais, um estudo realizado por [Gabrielsson e Lindström \(2001\)](#) e [Thayer e McNally \(1992\)](#) mostrou que existe uma relação entre essas características da música com os modelos dimensionais de emoção.

Dentre os elementos da música, a intensidade é uma característica básica que está altamente correlacionada com a excitação e é usada para classificar a dimensão da excitação nos modelos de referência para emoções ([ZHANG et al., 2017](#)). Outro elemento básico da música é o timbre que descreve a qualidade do som ([FU et al., 2011](#)) e é representado por algumas características chamadas de descritivas como coeficientes de *cepstrum* de *mel-frequência* (MFCCs), forma do espectro e contraste espectral (SC). Outra característica descritiva usada com frequência e que está associada ao timbre, é o histograma de coeficientes de wavelet Daubechies (DWCHs).

O ritmo é uma das características mais utilizadas no trabalho de ([FU et al., 2011](#)). Em sua pesquisa o autor descreve alguns dos padrões que mais apareceram e repetiram nos segmentos de música. Portanto, existe três características do ritmo que estão relacionadas com a resposta emocional humana: força do ritmo (*rhythm strength*), regularidade do ritmo (*rhythm regularity*) e tempo.

Todos esses elementos apresentados conceitualmente fazem parte da composição de um áudio de música. No entanto, esses elementos são características de nível de abstração da música também chamados de alto nível (*High-level*). Características como intensidade, timbre e ritmo, são refletidas por características descritivas (baixo nível ou *Low-level*) que podem ser extraídas diretamente dos arquivos de áudio, isso pode ser visto na Figura 3. Na Tabela 1 é apresentado algumas características descritivas que são comumente utilizadas para obter as características de alto nível da música.

**Tabela 1 – Características descritivas comumente usadas no MER**

ID	Característica	Sigla	Dimensão	Descrição
----	----------------	-------	----------	-----------

1	Mel-frequency Cepstrum Coefficients	MFCCs	20-D	Representação de baixa dimensão do espectro obtido a partir da distorção da escala Mel; reflete a sensibilidade de frequência não-linear do sistema auditivo humano
2	Octave-based Spectral Contrast	OSC	14-D	Representação áspera do conteúdo harmônico do domínio de frequência com base na identificação dos valores de pico e Vale do espectro de frequência.
3	Statistical Spectrum Descriptors	SSDs	4-D	Característica quádruplo composta do centróide espectral, fluxo, rolloff e planicidade, e geralmente relacionada à estrutura do timbre.
4	Chromagram	Chroma	12-D	Um método eficaz para a estimativa dos componentes do pitch em um curto intervalo de tempo; uma das principais propriedades é a captura de características harmônicas e melódicas da música, ao mesmo tempo em que são robustas as mudanças no timbre e na instrumentação.
5	Daubechies Wavelet Coefficient Histograms	DWCH	21-D	A média (7-D), o desvio padrão (7-D) e a energia sub-banda (7-D) do espectro.

Fonte: Adaptado de (YANG et al., 2017).

Além das características de músicas apresentadas acima podem também existirem outras características que levam ao reconhecimento da emoção musical. Por exemplo, informações de gênero e recursos líricos foram amplamente utilizados nos estudos relatados em (CHAPANERI et al., 2015) (CHIN et al., 2015) (LI; OGIHARA, 2003) (XUE et al., 2015) (YANG; LEE, 2004). O estudo de Li e Ogihara (2003) mostrou que a informação de gênero da música pode melhorar a precisão da detecção de emoção. Porém, quando uma música é classificada para um gênero errado, isso pode trazer um efeito negativo para os resultados do reconhecimento emocional (LAAR, 2006).

A etapa de selecionar quais as características são apropriadas é uma tarefa difícil, porque até mesmo os pesquisadores desconhecem quais recursos são mais importantes e úteis, além de ser difícil extrair os recursos perfeitos. Os pesquisadores então limitam as características e adequam ao seu escopo de trabalho de acordo com seu nível de conhecimento (YANG et al., 2017).

### 2.3.4 Métodos de Reconhecimento de Emoção Musical

Uma tarefa emocionante, mas também desafiadora dentro do estudo da música é construir um modelo computacional que possa reconhecer a emoção de uma música, além de organizar uma base de dados de músicas com apenas emoções descritas (KIM et al., 2010).

A parte essencial do modelo emocional é responsável pelo mapeamento entre as características descritivas do sinal de áudio para os recursos da música, como timbre, ritmo, melodia, entre outros. Por consequência esses recursos devem ser associados à emoção da música e realizar a anotação automática da emoção.

Katayose et al. (1988) foram os primeiros a conseguir a extração emocional de música em arquivos de símbolos em 1988 e Feng et al. (2003) foram os primeiros a configurar um sistema de recuperação de música baseada em emoção para arquivos de áudio em 2003. Também em 2003, Li e Ogihara (2003) relatou o problema de detecção de emoção como sendo um problema de classificação de vários rótulos pela primeira vez. Desde então, tem havido um número cada vez maior de resultados de pesquisa no campo de MER (YANG et al., 2017).

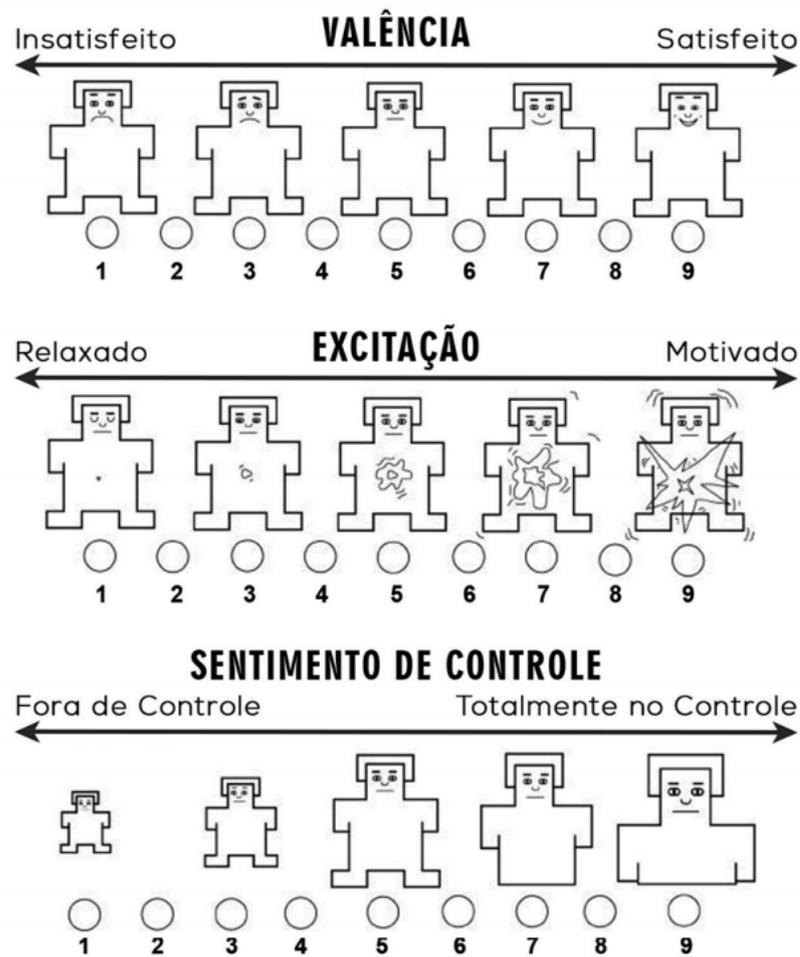
Com base nos diferentes métodos de modelagem, os métodos de MER podem ser classificados em duas categorias: classificação e regressão. Além do mais, devidas as diferentes combinações dos recursos de dados das músicas que existe durante a fase de modelagem, eles podem ser classificados em três categorias: aqueles que usam apenas os recursos musicais diferentes, aqueles que usam apenas os dados *ground-truth* e aqueles que usam a combinação destes dois recursos de dados.

## 2.4 Avaliação da Experiência do Usuário

Diversas abordagens têm sido utilizadas para avaliar a resposta emocional de usuários, por exemplo, o uso de instrumentos de autorrelato como Questionários, Think Aloud (SOMEREN et al., 1994), The Self-Assessment Manikin (SAM) (BRADLEY; LANG, 1994) e Emocards (REIJNEVELD et al., 2003). Além disso, outra alternativa para a avaliação emocional é o uso de sensores físicos para coleta de dados que podem ser modelados e analisados para se avaliar a resposta emocional do usuário.

No contexto desta pesquisa, o SAM foi escolhido porque este instrumento de autoavaliação pode avaliar os domínios que são diretamente correspondentes aos domínios da Estrutura do Espaço Emocional Semântico definido por (SCHERER, 2005), veja a Figura 4. A escala de avaliação é composto por três conjuntos de pictogramas que representam os domínios de Valência (satisfeito - insatisfeito), Excitação (motivado - relaxado) e Sentimento de controle (completamente no controle - fora de controle) (MAHLKE; MINGE, 2008). Cada domínio é representado por uma escala de valores que varia de 1-9. O usuário é instruído a escolher um dos cinco pictogramas ou intermediário entre os dois, o que melhor representa o seu estado

**Figura 4 – Cartão Self-Assessment Manikin (SAM) utilizado nos experimentos.**



**Fonte:** Adaptado de (BRADLEY; LANG, 1994).

emocional, no que diz respeito ao domínio que está sendo avaliado. Os valores dos círculos selecionados são verificados para obter o resultado. Para esta pesquisa foi realizado um mapeamento dos valores de 1 a 9 para uma escala entre -1 e 1, de maneira que pudessem corresponder a coordenadas no espaço semântico de Scherer.

# Capítulo 3

## TRABALHOS RELACIONADOS

---

---

### 3.1 Considerações Iniciais

Para condução deste trabalho foram elencados na literatura, trabalhos que tinham como objetivo próximo a automatização da interação entre um player mobile e um usuário, de forma que essa automatização pudesse ser feita por meio das emoções das músicas do usuário. Além disso, também foram verificados trabalhos que propuseram levar o usuário para um estado de emocional desejado utilizando alguma ferramenta.

A pesquisa foi realizada utilizando-se bases de dados para trabalhos científicos. As bases utilizadas para este trabalho foram: Periódicos da Capes<sup>1</sup>, SciELO<sup>2</sup>, Google Acadêmico<sup>3</sup>, IEEE Xplore<sup>4</sup>, além de Revistas e Livros.

Foram utilizados os seguintes termos como palavras chaves: music player emotion, recognition musical emotion, retrieval music emotion. Para filtrar os trabalhos encontrados, um mapeamento foi realizado de modo a encontrar os que relatavam sobre propôr uma automatização entre um player mobile e o usuário, por meio das emoções de suas músicas, ou levar o usuário para um estado emocional partindo do seu estado emocional atual.

Com isso, foram definidos oito trabalhos relacionados que apresentaram uma maior aproximação da proposta deste trabalho: *XPOD – A Human Activity and Emotion Aware Mobile Music Player* (DORNBUSH et al., 2005), que é apresentado na Seção 3.2, *HeartPlayer: A Smart Music Player Involving Emotion Recognition, Expression and Recommendation* (FAN et al., 2011) na Seção 3.3, *LyQ - An Emotion-aware Music Player* (HSU et al., 2006) apresentado na Seção 3.4, *Embedded Design of an Emotion-Aware Music Player* (CERVANTES; SONG, 2013) na Seção 3.5, *Facial Expression Based Music Player* (KAMBLE; KULKARNI, 2016) na Seção 3.6, *Emotion-based Music Recommendation Using Audio Features and User Playlist* (DENG; LEUNG, 2012) na Seção 3.7, *Emotion Based Mood Enhancing Music Recommendation* (IYER

<sup>1</sup> CAPES - [www.periodicos.capes.gov.br/](http://www.periodicos.capes.gov.br/)

<sup>2</sup> SciELO - [www.scielo.org/](http://www.scielo.org/)

<sup>3</sup> Google Acadêmico - [scholar.google.com.br](http://scholar.google.com.br)

<sup>4</sup> IEEE Xplore - <http://ieeexplore.ieee.org/>

et al., 2017) na Seção 3.8 e por fim na Seção 3.9 o trabalho *Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation* (GILDA et al., 2017).

## 3.2 XPOD – A Human Activity and Emotion Aware Mobile Music Player

No estudo realizado por Dornbush et al. (2005) - "*XPOD - A Human Activity and Emotion Aware Mobile Music Player*"- os autores consideraram a noção de colecionar informações de emoção e atividade humana do usuário para que assim pudesse explorar com essas informações uma possível evolução na experiência do usuário com os tocadores de música móvel. Para buscar alcançar tal objetivo foi adotado um dispositivo de leitor de mp3 móvel que automatiza o processo de seleção de música mais adequada à emoção e atividade atual do usuário.

O conceito do XPOD baseia-se na ideia de automatizar grande parte da interação entre o player de música e seu usuário. O dispositivo pode monitorar uma série de variáveis externas para determinar os níveis de atividade, movimento e estados físicos do usuário, para que assim possa fazer um modelo preciso de qual tarefa o usuário está realizando no momento atual e então prever o gênero de música apropriada para o usuário.

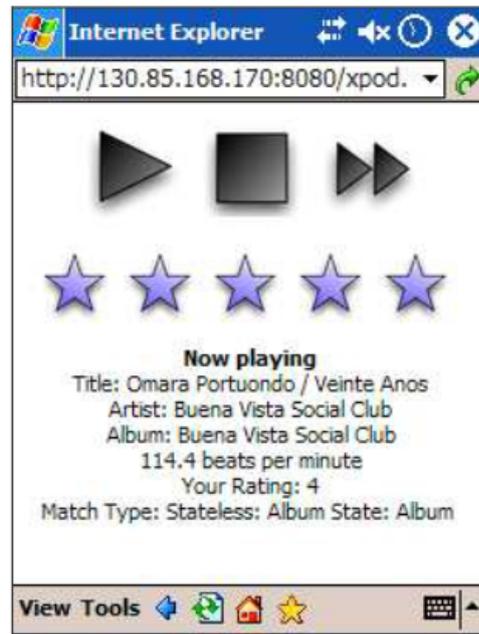
A princípio, o XPOD depende do usuário para treinar o player em avaliar a música preferida e em quais condições. Após um período de treinamento, o XPOD começa a usar seus algoritmos internos para fazer a seleção da música que melhor se adapte com a situação atual do usuário. O XPOD faz o uso de uma arquitetura Cliente/Servidor, onde todo o processamento é realizado na parte do servidor e as informações (lista de músicas e seus metadados) resultantes são enviadas para o dispositivo cliente XPOD. O Cliente é um dispositivo móvel PDA<sup>5</sup> (Personal Digital Assistant) e interface do XPOD pode ser vista na Figura 5.

Na interface do XPOD o usuário consegue visualizar as informações da música atual e decidirá se irá reproduzir ou não a música. Além da opção de reproduzir o usuário pode também avaliar a música que irá escutar. Uma vez que o usuário decide reproduzir a música, o servidor começa a receber um fluxo de informações através de sensores agregados ao usuário para determinar o estado atual.

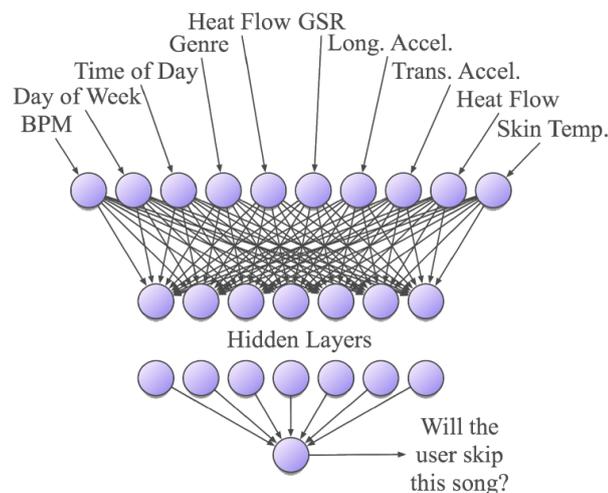
Quando o estado do usuário é determinado, essa informação é passada para o mecanismo de rede neural, que compara os níveis atuais de estado, tempo e atividade do usuário com as preferências anteriores da música correspondente ao conjunto de condições existente e faz uma seleção musical. A Figura 6 apresenta o formato da rede neural existente no XPOD.

Durante o estágio de treinamento do XPOD, o usuário escuta a música tocada e avalia-a por meio da interface no PDA. A preferência aplica a música como também em todos os seus

<sup>5</sup> PDA - Personal Digital Assistant, é um Assistente Digital Pessoal que geralmente é conhecido como: Palm, Palmtop, WinCE, PocketPC, Handheld, entre outros.

**Figura 5 – XPOD interface do usuário.**

Fonte: (DORNBUSH et al., 2005).

**Figura 6 – XPOD Formato da rede neural.**

Fonte: (DORNBUSH et al., 2005).

atributos (artista, gênero, etc). Ao comparar estados do corpo do usuário passado, BPM, gêneros e autores das opções de músicas passadas para o conjunto existente de variáveis, a rede neural toma uma decisão informada e seleciona uma música apropriada.

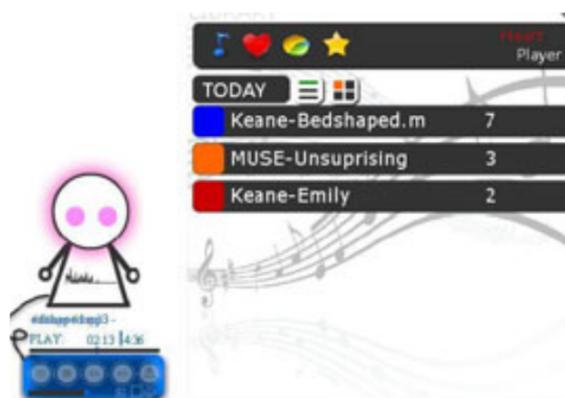
### 3.3 HeartPlayer: A Smart Music Player Involving Emotion Recognition, Expression and Recommendation

No trabalho "*HeartPlayer: A Smart Music Player Involving Emotion Recognition, Expression and Recommendation*" Fan et al. (2011) propõem um *player* de música inteligente que

ajuda o usuário a "ver" a música de uma forma visual para que o usuário possa entender e sentir a emoção da música. Além de analisar as músicas que o usuário ouve o HeartPlayer identifica o humor do usuário para recomendar músicas mais adequadas.

O HeartPlayer é composto por quatro partes: o player, uma figura, a lista de músicas do usuário e programa que efetua o cálculo das emoções. A interface do HeartPlayer pode ser vista na Figura 7.

**Figura 7 – Interface HeartPlayer.**



**Fonte:** (FAN et al., 2011).

A Figura 7 apresenta a tela principal do player e o conjunto de músicas do usuário. Ao lado do conjunto das músicas possui a imagem um personagem criado pelos autores que é responsável por mostrar e expressar a compreensão da emoção da música. Assim que a emoção é calculada, a figura muda suas expressões faciais. Os movimentos são alterados de acordo com a informação instantânea da música, por exemplo, com o uso da característica de amplitude e do espectro da música.

Na biblioteca de músicas do player apresentado, há quatro seções: "Meus Favoritos", "Humor de Hoje", "Registro de Emoções" e "Personalidade da Música" as quais são apresentadas em forma de gráficos, veja o exemplo na Figura 8

**Figura 8 – HeartPlayer em execução.**



**Fonte:** (FAN et al., 2011).

A Figura 8 apresenta dois exemplos citados e portanto, utilizam seis cores para representar seis emoções musicais. Assim com os resultados da análise a recomendação pode ser feita.

Há uma parte específica do HeartPlayer que é responsável pelo cálculo da emoção das músicas. Os autores decidiram recuperar alguns recursos da música, como: tempo, amplitude e melodia. Dessa forma construíram um modelo que utilizassem desses recursos para efetuar o cálculo da emoção musical.

São definidos seis intervalos de pontuação que correspondem às seis emoções do Heart-Player, são elas: "excitadas", "alegres", "pacíficas", "dolorosas", "desesperadas" e "nervosas". O programa calcula a pontuação e, se a pontuação cair em um dos intervalos, a emoção da música é identificada.

### 3.4 LyQ - An Emotion-aware Music Player

No trabalho de Hsu et al. (2006) - "*LyQ - An Emotion-aware Music Player*" - o autor apresenta uma proposta ampla para o player LyQ. O LyQ é projetado para explorar o senso de afeto para a detecção de músicas com base em teorias psicológicas, semânticas, senso comum, bem como a experiência pessoal humana com música.

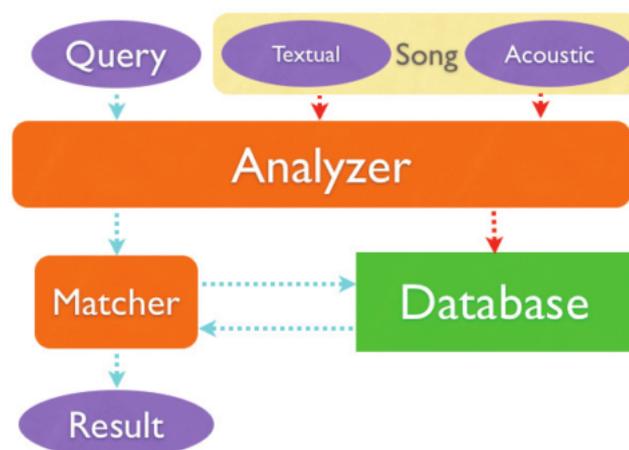
Para o autor, uma música contém várias características que podem estar relacionadas com a sua detecção de emoção. Dentre essas características a música pode conter seu estado emocional para sua letra, como também seu acústico e seus metadados associados (título, artista, álbum, data e opiniões) (HSU et al., 2006).

Uma das características projetadas pelo LyQ é a capacidade de extrair a emoção não apenas do acústico das músicas, mas como também do senso comum que existe na sua letra musical. O autor define que utiliza as capacidades de sensor de efeitos da ConceptNet na criação desta aplicação. A ConceptNet é uma base de conhecimento de senso comum e um conjunto de ferramentas de processamento de linguagem natural que geram automaticamente uma rede de senso comum usando as declarações de senso comum no banco de dados OMCS (Oracle Management Cloud). Este programa oferece muitas funções úteis, tais como tópicos, percepção de afetos, criação de analogias, expansão contextual, entre outras (LIU; SINGH, 2004).

O LyQ foi projetado para receber três entrada: *Query* (consulta), *Textual* (conteúdo textual) e o *Acoustic* (conteúdo acústico), logo sua arquitetura pode ser vista na Figura 9.

Uma das entradas do LyQ é a consulta descrita em linguagem natural dada pelo usuário. O usuário pode inserir uma frase que expresse sua sensação, por exemplo, "Estou preparando uma festa hoje a noite!", de forma que expresse a emoção e o desejo do usuário. A outra entrada é por meio de uma música ou gravação de som realizada. A música ou gravação é subdividida em duas novas entradas, a parte textual do áudio e sua parte acústica.

O LyQ Analyzer adota diferentes modelos de avaliação com base no tipo da entrada.

**Figura 9 – LyQ: Arquitetura.**

**Fonte:** (HSU et al., 2006).

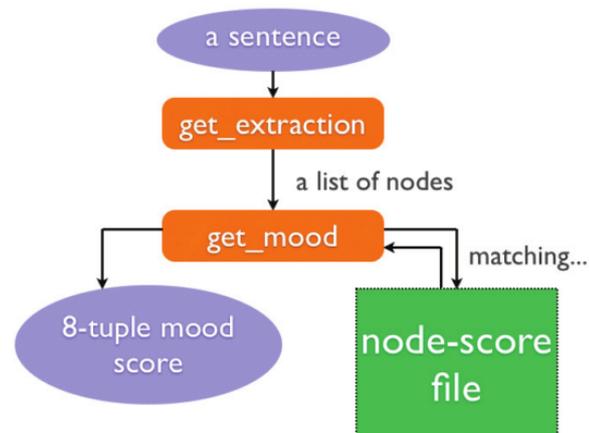
Esses recursos extraídos são armazenados no banco de dados. De acordo com a consulta do usuário, o Matcher procura as cinco canções com sensação implícita e significado semântico como a sensibilidade da consulta. Finalmente, o LyQ recomenda as músicas correspondentes ao usuário.

Os autores afirmam que a parte principal do analisador em relação à análise textual é o seu modelo de detecção de emoção. Para construir seu modelo de percepção de emoções é utilizada uma tupla de oito para representar as dimensões do humor da música, que são: irritado, emocionante, feliz, brincalhão, relaxado, triste, espiritual e frágil.

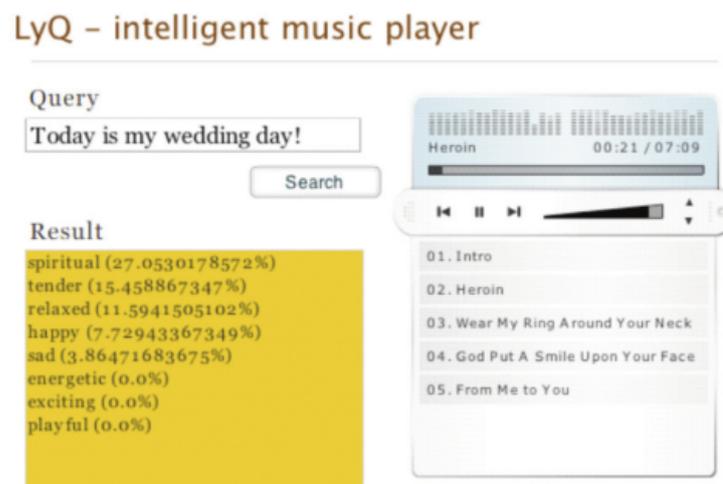
A sequência de como é feita a recuperação da emoção no LyQ pode ser vista na Figura 10. Uma frase é dada e não importa o texto da consulta, seja um título de uma música ou uma frase aleatória. O LyQ obtém um objeto de extração que contém um resumo analisado do texto a partir da função *get\_extraction* da ConceptNet. Em seguida é combinado com o mesmo termo que ocorre no arquivo de nó de pontuação (*node-score-file*), assim é retornado o valor emocional da consulta.

De forma geral, o protótipo da interface projetada do LyQ é apresentada na Figura 11. Em sua interface pode ser vista três partes distintas, embora estejam correlacionadas. A primeira parte é a *Query*, local em que o usuário irá inserir sua frase de consulta em linguagem natural. Na segunda parte é possível visualizar todos os valores de humores referentes a entrada que foi inserida pelo usuário expressado em porcentagem.

Por fim, a terceira é o reproduzidor da música que está localizada à direita na Figura 11. Ao recuperar a emoção expressa pela entrada do usuário, o LyQ recomenda pelo menos cinco músicas no leitor já pré classificadas de acordo com a emoção obtida.

**Figura 10 – LyQ: Diagrama de fluxo para recuperação de recursos de emoção.**

Fonte: (HSU et al., 2006).

**Figura 11 – LyQ: Interface.**

Fonte: (HSU et al., 2006).

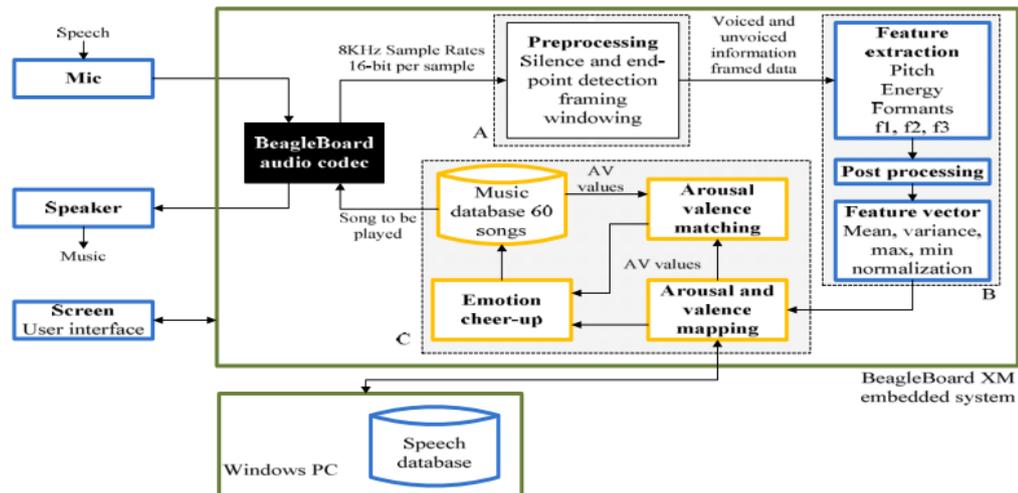
### 3.5 Embedded Design of an Emotion-Aware Music Player

O trabalho "*Embedded Design of an Emotion-Aware Music Player*" (CERVANTES; SONG, 2013) busca encontrar uma solução possível para relacionar a música e seu conteúdo emocional utilizando categorias emocionais contínuas. É utilizado um modelo emocional contínuo que permite uma melhor diferenciação de músicas que possuam seu conteúdo emocional semelhante, mas que ainda são diferentes entre si.

O projeto dos autores buscam contribuir para área de interação humano-robô (HRI - *human-robot interaction*), propondo uma arquitetura para um sistema eletrônico que reproduza músicas com base a emoção do ser humano. Para isso o sinal da fala é utilizado para criar um player de música com base na emoção da fala. O sinal da fala é mapeado para um plano bidimensional de valência e excitação. Com isso, o sistema busca automaticamente no banco de

dados músicas as quais também são expressas em valores de valência e excitação. Outra proposta do trabalho é que as músicas são escolhidas para que possam levar o usuário para um estado mais neutro/feliz.

**Figura 12 – Diagrama de blocos da arquitetura do sistema emotion-aware music player.**



Fonte: (CERVANTES; SONG, 2013).

A Figura 12 mostra a arquitetura do sistema emotion-aware music player descrita no trabalho de Cervantes e Song (2013). O sistema utiliza uma plataforma embarcada chamada BeagleBoard XM, acopladas com um microfone de captura de áudio, uma caixa externa para reprodução das músicas, além de um monitor *touchscreen*<sup>6</sup>.

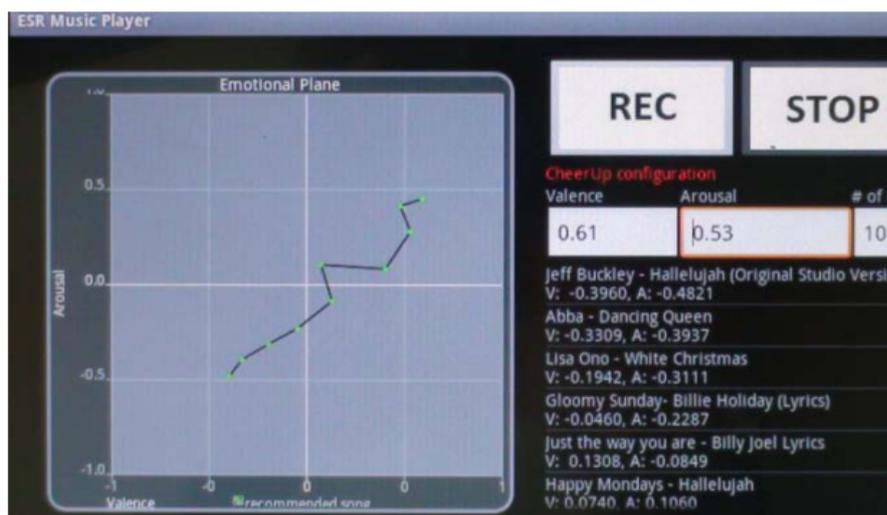
O sistema possui três blocos principais: pré-processamento de sinal, extração de recursos e mapeamento de excitação e valência. No bloco de pré-processamento de sinal e extração de recursos funcionam para obter informações úteis do sinal da fala, nessas etapas são removidos qualquer silêncio e ruído encontrados no sinal. Para esse trabalho os autores decidiram recuperar as seguintes características descritivas do áudio da fala: *Energy*, *Pitch* e *Formants*.

Após obter os recursos do áudio da fala é construído o vetor de características com as informações normalizadas para que assim possa ser mapeado em um espaço emocional contínuo. O autor justifica que o uso desse mapeamento é devido à facilidade em relacionar a emoção da fala com a emoção da música (CERVANTES; SONG, 2013).

Através da Figura 13 é possível visualizar a tela do player de música proposto pelos autores deste trabalho. Percebe-se que o usuário possui a opção de gravar sua fala por um determinado tempo, assim em um gráfico bidimensional é exposto os valores de valência e excitação no decorrer da gravação da fala.

<sup>6</sup> *Touchscreen*: tela sensível ao toque.

Figura 13 – Captura da tela do player de música.



Fonte: (CERVANTES; SONG, 2013).

Além das informações da fala é possível visualizar as informações das músicas que são reproduzidas automaticamente, por exemplo, título, autor e valores de valência e excitação. As músicas são selecionadas a partir de um banco de dados de 60 músicas do próprio sistema. Entretanto, essas músicas já se encontram classificadas com seus valores de excitação e valência.

### 3.6 Facial Expression Based Music Player

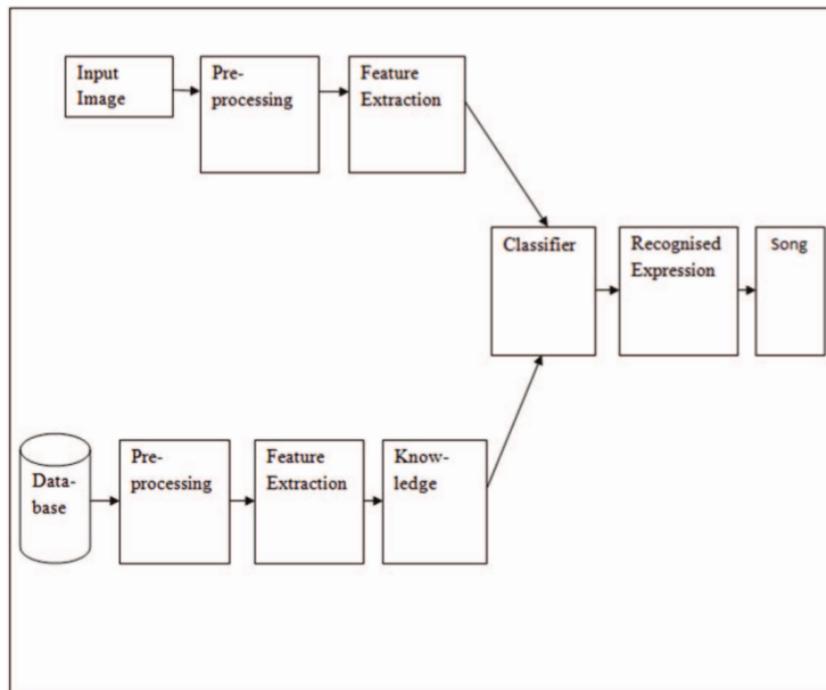
Em "*Facial Expression Based Music Player*" os autores [Kamble e Kulkarni \(2016\)](#) visam desenvolver um sistema inteligente que possa facilmente reconhecer a expressão facial e, consequentemente, reproduzir uma música baseada com a emoção obtida. Nesse trabalho é utilizado sete classes de emoções: tristeza, raiva, desgosto, medo, surpresa e neutro.

A arquitetura do sistema proposto por [Kamble e Kulkarni \(2016\)](#) pode ser observada na Figura 14. O sistema recebe como entrada uma imagem no formato .jpg, então a imagem é passada para uma etapa de pré-processamento. Nessa etapa é realizada a remoção de informações indesejadas, como cor de fundo, iluminação, e redimensionamento. Em seguida os recursos necessários são extraídos da imagem e armazenados como informações úteis.

Posteriormente, esses recursos são adicionados ao classificador, que por meio do cálculo da distância euclidiana reconhece a expressão. Por fim, uma faixa de música será tocada com base a emoção obtida.

De forma geral, todo o sistema foi projetado em 5 etapas:

1. **Aquisição da imagem:** as imagens podem ser adquiridas através de alguma câmera ou imagens de base de dados online, desde que estejam no formato .jpg.

**Figura 14 – Arquitetura do Sistema.**

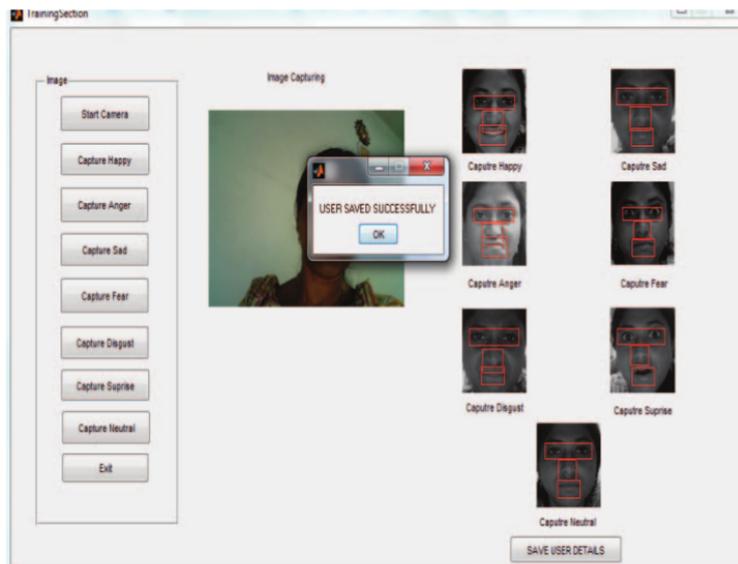
Fonte: (KAMBLE; KULKARNI, 2016).

2. **Pre-processamento:** essa etapa busca eliminar informações indesejadas. As imagens são convertidas de RGB (Red, Green, Blue) para escala de cinza e são redimensionadas para 256 \* 256 pixels. Durante o pré-processamento, os olhos, nariz e boca são considerados a região de interesse.
3. **Extração de recursos faciais:** Após o pré-processamento, a próxima etapa é a extração de recursos. Nessa etapa as características faciais extraídas são armazenadas como a informação útil sob a forma de vetores durante a fase de treinamento e fase de teste. As seguintes características faciais são consideradas "Boca, testa, olhos, tez da pele, coxas do rosto e do queixo, sobrancelhas, nariz e rugas no rosto".
4. **Reconhecimento de expressões:** O classificador de distância euclidiana é utilizado para reconhecer e classificar as expressões de uma pessoa. Obtém a correspondência mais próxima para os dados de teste do conjunto de dados de treinamento e dá uma melhor combinação para a expressão atual detectada.
5. **Reprodução da música:** é a etapa mais importante do sistema, segundo os autores. É o passo em que a música é reproduzida com base na emoção atual detectada de um indivíduo. Uma vez que a expressão fácil do usuário é classificada, o estado emocional correspondente é reconhecido, assim uma série de músicas pertencentes a emoção detectada é colocada na lista para reprodução

A Figura 15 mostra a fase de treinamento em que o usuário posa para todas as 7 expressões

definidas para o sistema. Além dessa parte do sistema, há outra que é responsável pelos testes, as expressões são reconhecidas e então exibida.

**Figura 15 – Captura da tela de treinamento do sistema.**



Fonte: (KAMBLE; KULKARNI, 2016)

### 3.7 Emotion-based Music Recommendation Using Audio Features and User Playlist

Deng e Leung (2012) propuseram um sistema que realiza recomendações de músicas por emoção usando recursos de áudio e listas de reprodução dos usuários. O trabalho tem como principal contribuição a proposta de analisar a lista de reprodução dos usuários, rastrear a emoção de cada música, e então recomendar músicas que não apenas se ajustam melhor às suas emoções, mas também a adoção de abordagens baseadas em conteúdo.

A emoção musical é representada utilizando a representação de emoção dimensional chamada Ressonância-excitação-Valência (RAV) e para recomendar as músicas com base na emoção do usuário previsto, foi escolhido uma lista de classificação otimizada que apresenta as maiores similaridades emocionais da música com o estado de emoção previsto.

Os autores propuseram a recomendação de música com base na emoção, transmitindo na sessão da lista de reprodução histórica do ouvinte. Através da lista de reprodução do ouvinte em uma sessão, obtém-se implicitamente uma sequência de estado de emoção correspondente que tem grandes efeitos sobre o próximo estado de emoção. Com isso, são capazes de calcular a influência da emoção de cada peça de música na sequência para o estado de emoção atual, respectivamente. Para isso, utilizam a teoria da probabilidade discriminativa e condicional para obter a saída da probabilidade do estado atual da emoção prevista, dadas as observações.

O método de iteração de minimização é utilizado para computar a lista de músicas recomendadas com classificação otimizada. Os resultados mostraram que a abordagem proposta é eficaz para recomendações personalizadas de música baseada em emoções.

### 3.8 Emotion Based Mood Enhancing Music Recommendation

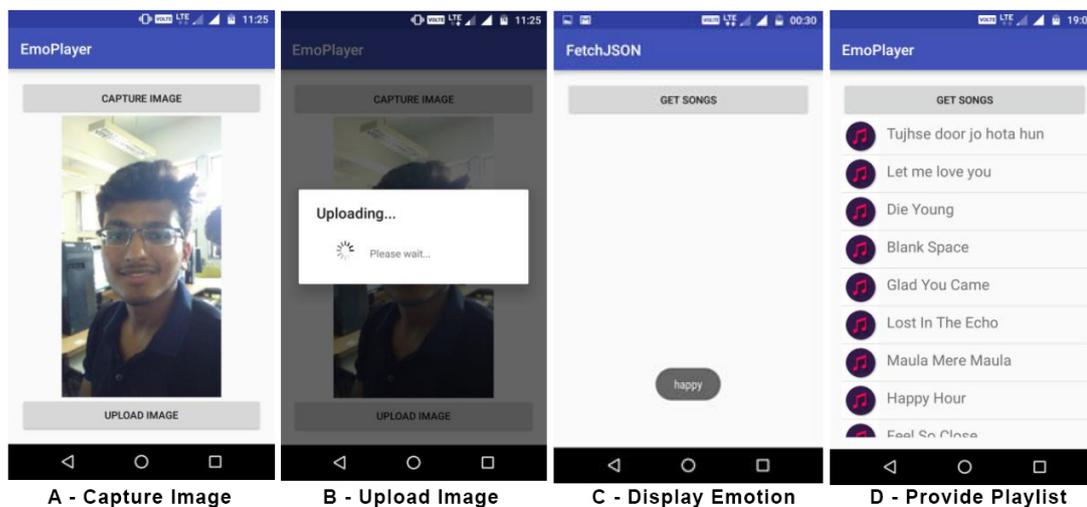
Iyer et al. (2017) em seu trabalho "*Emotion Based Mood Enhancing Music Recommendation*" propuseram um aplicativo para sistema Android que pode minimizar os esforços do usuário em classificar sua lista de músicas baseadas na emoção. O sistema captura a imagem do rosto do usuário e em seguida ele detecta a emoção e faz uma lista de músicas que irão melhorar seu humor enquanto as músicas são tocadas.

No artigo o autor, propõe o sistema "EmoPlayer", que é um aplicativo baseado em Android. O aplicativo identifica a emoção do rosto e com base na emoção reconhecida, ele envia a emoção para o servidor de música e buscará uma lista de reprodução adequada.

Assim que o usuário inicia o aplicativo, ele verifica se há conectividade com a Internet. Se o usuário não conseguir se conectar à Internet, o aplicativo exibirá as músicas da memória do dispositivo. No entanto, no estabelecimento de uma internet estável, o usuário tem a opção de tocar músicas com base em sua emoção ou diretamente do servidor de música. Uma vez que ele opte por uma recomendação de música baseada em emoção, o aplicativo inicia a câmera do dispositivo e o usuário tem que capturar sua imagem. Esta imagem é enviada para o servidor de imagens. A detecção de rosto é então realizada no servidor.

O próximo passo é o reconhecimento de emoções. O servidor de imagem envia a emoção detectada para o dispositivo, exibe para o usuário e envia emoção para o servidor de música. O servidor de música recebe essa emoção e usa um código que sugere ao usuário uma lista de músicas baseadas nessa emoção. As músicas serão sugeridas de tal forma que as poucas músicas iniciais refletirão a emoção atual e as músicas na lista serão de uma emoção mais feliz. A Figura 16 apresenta as telas da interface do aplicativo EmoPlayer.

O autor conclui que apesar de o sistema proposto ser capaz de processar a imagem facial e reconhecer as emoções básicas e depois reproduzir música com base nessas emoções, espera-se que para trabalhos futuros a aplicação possa concentrar em melhorar a taxa de reconhecimento de emoções e reconhecer um número maior de emoções.

**Figura 16 – Interface EmoPlayer.**

Fonte: (IYER et al., 2017).

### 3.9 Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation

Gilda et al. (2017) apresentaram um reprodutor de música multiplataforma chamada de EMP, que pode recomendar músicas em tempo real com base na emoção do usuário. O EMP realiza as recomendações de músicas baseada nas emoções, incorporando as capacidades do raciocínio de contexto. O reprodutor é dividido em três módulos: Módulo de Emoção, Módulo de Classificação de Música e Módulo de Recomendação. O Módulo de Emoção pega uma imagem do rosto do usuário como uma entrada e usa algoritmos de aprendizagem para identificar sua emoção. O Módulo de Classificação de Música classifica músicas em 4 diferentes classes de humor. O Módulo de Recomendação sugere músicas para o usuário, mapeando suas emoções para o humor da música, considerando as preferências do usuário.

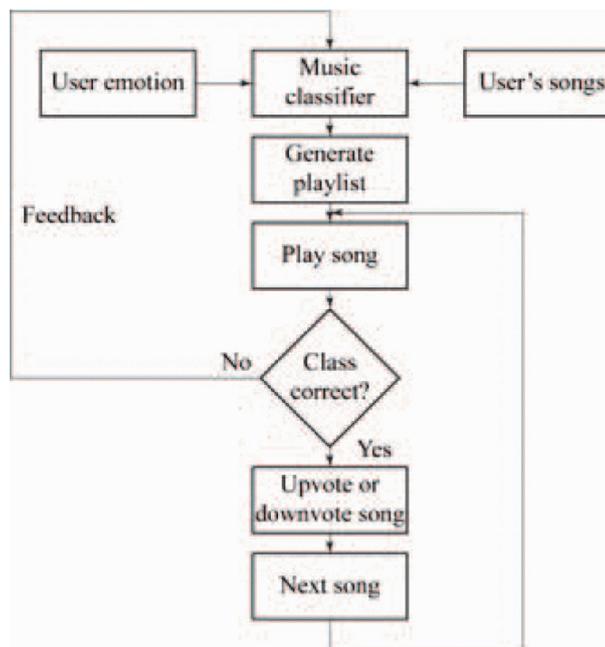
O módulo de emoção utiliza redes neurais convolucionais (CNNs) para construir um modelo computacional que classifica com sucesso a emoção em quatro modos: feliz, triste, zangado e neutro. A rede final foi treinada em 20.973 imagens e testada em 5.244 imagens. No final, o modelo alcançou uma precisão de 90,23%.

No módulo de classificação os recursos acústicos das músicas foram extraídos utilizando a biblioteca LibROSA. Com base nesses recursos, uma rede neural artificial que classifica com sucesso as músicas em 4 classes foi treinada com uma precisão de 97,69%. Os recursos do processo de extração pertenciam a diferentes classes: espectral (RMSE, centróide, rolloff, MFCC, curtose, etc.), rítmica (tempo, espectro de batida, etc.), modo tonal e tom. Uma precisão de 97,69% foi obtida da classificação geral.

O módulo de recomendação é responsável por gerar uma lista de reprodução de músicas para o usuário. Ele permite que o usuário modifique a lista de reprodução com base em suas

preferências e modifique os rótulos de classe das músicas também. A Figura 17 mostra o funcionamento do módulo de recomendação.

**Figura 17 – Módulo de recomendação.**



Fonte: (GILDA et al., 2017).

As músicas classificadas são mapeadas para o humor do usuário. Após a conclusão do procedimento de mapeamento, é gerada uma lista de reprodução de músicas relevantes. Músicas semelhantes são agrupadas ao gerar a lista de reprodução. O mecanismo de recomendação possui um mecanismo duplo que recomenda músicas baseadas no humor percebido pelo usuário e suas preferências.

Os autores ainda implementaram um player de música adaptável, que utiliza o algoritmo de aprendizado de máquina SGD (Descentramento Estocástico de Gradiente). Se o usuário quiser alterar a classe de uma música em particular, o SGD é implementado considerando o novo rótulo apenas para esse usuário específico.

Os resultados dos autores mostram que a alta precisão e o tempo de resposta rápido do aplicativo o tornam adequado para a maioria dos propósitos práticos. O módulo de classificação de música em particular, executa significativamente bem; alcança alta precisão na categoria “raiva”, ao mesmo tempo em que atua sensivelmente bem nas categorias “feliz” e “calma”. Assim, o EMP reduz os esforços do usuário para gerar listas de reprodução, mapeando eficientemente a emoção do usuário para a classe de música correta com uma precisão geral de 97,69%.

### 3.10 Comparação dos Trabalhos Relacionados

No trabalho de Dornbush et al. (2005) é apresentado o player XPOD que tem como objetivo automatizar a interação do usuário com o player e para isso faz o uso de sensores para

obter a informações fisiológicas do usuário. No entanto, essas informações não são utilizadas para determinar o estado emocional do usuário, mas sim indicar qual música tocar com base algum estado passado em que o usuário se encontrava. Todo o sistema se concentra em armazenar as informações do usuário quando se ouve uma determinada música. Assim, o player busca indicar uma música que está nas mesmas condições ou próximas ao estado atual do usuário.

Além do mais, o trabalho não realiza nenhum processamento de sinal do áudio da música para recuperar a emoção e indicar ao usuário. Para os autores, o XPOD é uma implementação inicial que fornece a base para uma maior exploração de sistemas móveis conscientes de humanos e emoções (DORNBUSH et al., 2005).

Assim como o player XPOD, o player HeartPlayer proposto por Fan et al. (2011) não obtém a emoção do usuário a partir de uma resposta fisiológica ou sentimento subjetivo. Nesse trabalho a emoção do usuário é detectada a partir do histórico de preferências de músicas ouvidas pelo usuário. Em contrapartida, e diferentemente dos trabalhos de Dornbush et al. (2005), Cervantes e Song (2013) e Kamble e Kulkarni (2016) que indica as músicas a partir de uma base dados com músicas já classificadas, o HeartPlayer obtém a emoção das músicas realizando o processamento de cada uma. Com isso, o player pode expressar as músicas do usuário através da sua emoção obtida.

O LyQ é um player proposto pelos autores Hsu et al. (2006) que além de obter a emoção das músicas do usuário permite também a utilização de uma base de dados online para recomendação de músicas caso o usuário não configure o caminho em que se encontra suas músicas. O estágio em que o LyQ se encontra não permite obter a emoção do usuário a partir de alguma resposta fisiológica ou sentimento subjetivo. A emoção é detectada por meio da análise textual da entrada informada pelo usuário. Com isso a música é recomendada por um banco online ou através das músicas do usuário em um computador.

Outro trabalho que não realiza o processamento do sinal de áudio da música para obter a emoção é o trabalho "*Embedded design of an emotion-aware music player*" dos autores Cervantes e Song (2013). Nesse trabalho há o processamento apenas do áudio da fala do usuário através do uso de um sensor de captura de áudio. A emoção do usuário é obtida e as músicas são recomendadas ao usuário através de um banco de dados com 60 músicas já classificadas. Diferente dos demais trabalhos o Emotion-Aware é um player desenvolvido para dispositivos embarcados que ao detectar a emoção do usuário o player busca na base de dados por músicas que levam o usuário para um estado mais alegre ou neutro.

Assim como no trabalho de Cervantes e Song (2013), os autores Kamble e Kulkarni (2016) propuseram um player que reproduz músicas de acordo com a emoção do usuário e para alcançar esse objetivo ao invés de utilizar um sensor de captura de áudio os autores utilizaram uma câmera para detectar a emoção atual do usuário. Além disso, os autores não se preocuparam em utilizar a emoção obtida do usuário para recomendar músicas que levassem o usuário para um estado mais alegre ou neutro. Os autores também utilizaram uma base de dados com músicas

já classificadas e as recomendou com base no estado atual do usuário.

Os três últimos trabalhos apresentados possuem como proposta principal a recomendação de músicas para ajustar a emoção do usuário. O trabalho de [Deng e Leung \(2012\)](#) utiliza as listas de música do usuário para recomendar músicas que ajustam a sua emoção atual. No entanto, o usuário poderia estar em um estado emocional "Depressivo", e com isso músicas depressivas seriam retornadas para a reprodução.

Já no trabalho de [Iyer et al. \(2017\)](#), as músicas são recomendadas a partir da lista do usuário, com o objetivo de melhorar o estado emocional do usuário. Então, se o usuário estiver em um estado emocional "Depressivo", as músicas de sua própria preferência serão retornadas, de modo que poderão melhorar o seu estado emocional. No entanto, o usuário pode se encontrar em um estado de alta excitação, buscando músicas que o leve para um estado "Apaixonado" ou "Solene". Então, o aplicativo para Android proposto não seria capaz de recomendar músicas que o levasse para um estado emocional desejado.

Por fim o trabalho de [Gilda et al. \(2017\)](#), assim como o trabalho de [Deng e Leung \(2012\)](#), também recomendam músicas que ajustam emoção atual do usuário, ou seja, se o usuário estiver depressivo, então músicas depressivas também serão retornadas. Além do mais, as músicas do usuário são classificadas em apenas 4 classes de emoções discretas.

A Tabela 2 apresenta um comparativo entre os trabalhos relacionados que mais se aproximam com o objetivo desse trabalho.

**Tabela 2 – Comparativo de trabalhos relacionados.**

Trabalhos	Obtém Emoção da Música	Espaço Semântico Dimensional	Lista do Usuário	Emoção Desejada	Reprodutor	Mobile
XPOD	NÃO	NÃO	SIM	NÃO	SIM	SIM
HeartPlayer	SIM	NÃO	SIM	NÃO	SIM	NÃO
LyQ	SIM	NÃO	SIM	SIM	SIM	NÃO
Emotion-Aware	SIM	SIM	NÃO	NÃO	SIM	NÃO
Facial Expression	SIM	NÃO	NÃO	NÃO	SIM	NÃO
Deng and Leung	SIM	SIM	SIM	NÃO	NÃO	NÃO
Iyer et al.	SIM	NÃO	SIM	NÃO	SIM	SIM
Gilda et al. EMP	SIM	NÃO	SIM	NÃO	SIM	SIM
Trabalho Proposto	SIM	SIM	SIM	SIM	SIM	SIM

**Fonte:** Elaborada pelo autor.

É apresentada cinco características na tabela, sendo elas:

1. **Obtém Emoção da Música:** essa característica representa trabalhos que realizam um processamento no áudio da música para extrair emoção e de alguma forma recomendá-las sem a utilização de um banco de dados com músicas classificadas.

2. **Espaço Semântico Dimensional:** trabalhos que classificam as emoções em um modelo dimensional, diferente de trabalhos que mapeiam as emoções para rótulos.
3. **Lista do Usuário:** característica que representa trabalhos em que os autores analisam as próprias músicas do usuário, ou seja, considera o estilo musical de preferência do ouvinte.
4. **Emoção Desejada:** trabalhos que consideram a emoção em que o usuário deseja alcançar e não apenas a emoção atual.
5. **Reprodutor:** são trabalhos que propõem um reprodutor de música.
6. **Mobile:** trabalhos que desenvolveram um reprodutor de música para dispositivos móveis.

Este trabalho de mestrado abrange as seis características apresentadas que foram utilizadas para comparação dos trabalhos elencados da literatura. A conclusão deste trabalho produziu a ferramenta m-Motion, que diferente dos trabalhos relacionados, busca ajudar o usuário alcançar um estado emocional desejado, partindo do seu estado emocional atual. Muitos trabalhos na literatura, assim como os trabalhos relacionados apresentados nas seções anteriores, os autores verificam o estado emocional atual do usuário e recomendam músicas para o estado detectado. Verificar apenas o estado emocional atual para recomendar, pode se sujeitar em uma recomendação que levará o usuário para um estado emocional de baixa excitação, caso seu estado emocional atual seja de baixa excitação. O m-Motion possui cautela nas recomendações e utiliza o conjunto de músicas de preferência do usuário, além de considerar o estado emocional desejado.

Os trabalhos relacionados apresentados, com exceção o trabalho de [Dornbush et al. \(2005\)](#), os autores recuperam a emoção das músicas e mapeiam para modelos que descrevem as emoções em espaços dimensionais ou por rótulos. No m-Motion as emoções são descritas utilizando um espaço dimensional, diferente dos trabalhos de [Dornbush et al. \(2005\)](#), [Fan et al. \(2011\)](#), [Hsu et al. \(2006\)](#), [Kamble e Kulkarni \(2016\)](#), [Iyer et al. \(2017\)](#) e [Gilda et al. \(2017\)](#) que utilizam um conjunto limitado de classes de emoções. Além do mais, o m-Motion é um reprodutor de música móvel para *smartphones*<sup>7</sup> que se faz disponível e acessível para os usuários devido ao grande avanço das tecnologias móveis.

---

<sup>7</sup> Smartphone - Palavra inglesa que significa "telefone inteligente" é um celular.

# Capítulo 4

## M-MOTION

---

---

### 4.1 Considerações Iniciais

Este trabalho de mestrado tem como objetivo ajudar o usuário a alcançar um estado emocional desejado partindo do seu estado emocional atual por meio da seleção automática de músicas. Para alcançar esse objetivo, uma ferramenta de reprodução de música que adapta as músicas do usuário pela emoção foi construída para dispositivos móveis (*smartphones*). A princípio, a emoção do usuário é recuperada por meio de sentimentos subjetivos, em seguida o reprodutor de música recupera a emoção das músicas do usuário e então uma nova lista de músicas é organizada para que o usuário atinja seu estado emocional desejado.

Uma das maneiras de classificar as emoções surgiram com o trabalho desenvolvido por [Scherer \(2005\)](#), denominado estrutura do espaço emocional semântico. O espaço semântico está dividido em oitantes, com domínios positivos e negativos, de acordo com a excitação, sentimento de controle, valência e facilidade de conclusão de objetivo. Para definir a emoção relacionada à música e do usuário, este trabalho irá utilizar o espaço semântico desenvolvido por [Scherer \(2005\)](#).

No entanto, para este trabalho utilizamos duas dimensões do modelo de Scherer, sendo excitação e valência. A escolha dessas dimensões se deu pelo fato do *dataset* definido para este trabalho conter estimações para essas duas dimensões. Para cada música do usuário é extraído um conjunto de recursos de áudio para prever os valores de valência e excitação. Em seguida esses valores são mapeados para o espaço semântico de Scherer e então definido a emoção da música. A emoção do usuário que é obtida por meio de sentimentos subjetivos também é mapeada para o espaço semântico de Scherer.

### 4.2 Arquitetura

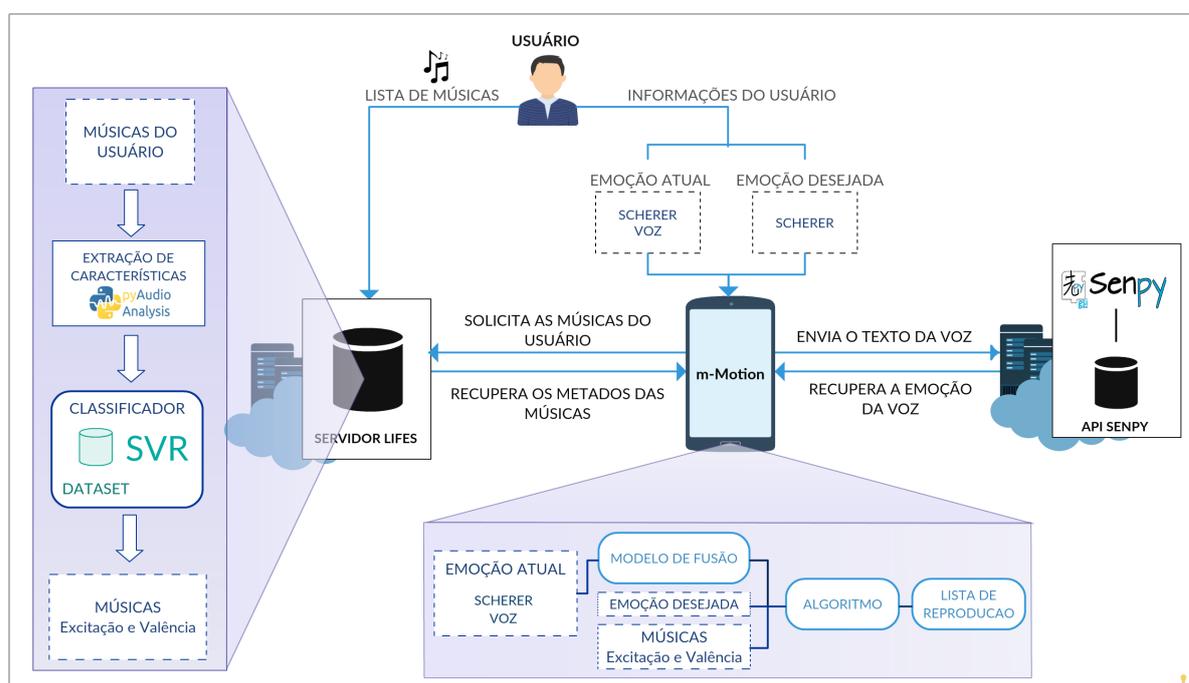
Com o rápido desenvolvimento de smartphones e o avanço da Internet, cada vez mais pessoas passaram a preferir ouvir músicas pela internet, pois podem escolher suas músicas de

acordo com suas preferências (WANG, 2007). Na última década, as novas tecnologias abriram caminho para novos canais de distribuição de conteúdo digital, por exemplo, plataformas de transmissão de streaming de músicas como Spotify<sup>1</sup> ou Apple Music<sup>2</sup>. Ao mesmo tempo, dispositivos móveis como smartphones ou tablets permitem que seus usuários acessem milhões de faixas nas plataformas de transmissão em várias situações durante todo o dia (PICHL et al., 2017).

Esses desenvolvimentos tornaram a organização da música um tópico altamente interessante e, com isso, fez com que a recomendação de música tornasse sensível ao contexto, pois os usuários têm o desafio de encontrar uma música que eles gostam em uma ampla variedade de música oferecida pelas plataformas de transmissão de música. Em princípio, os usuários precisam navegar através de sua coleção de música para encontrar a música que eles pretendem ouvir durante diferentes atividades ou situações (YANG; HE, 2017).

Diante desse avanço de plataformas de streaming de músicas bem como o aumento de usuários que possui a preferência de ouvir música pela internet, este trabalho adotou o modelo semelhante ao de um streaming de música convencional e a arquitetura proposta deste trabalho pode ser vista na Figura 18.

**Figura 18 – Arquitetura m-Motion.**



**Fonte:** Elaborada pelo autor.

De maneira geral, as plataformas de streaming armazenam as músicas dos usuários em seus servidores. No entanto, este trabalho buscou evitar qualquer tipo de violação de direitos autorais sobre o armazenamento e reprodução das músicas por meio do aplicativo m-Motion. As

<sup>1</sup> <http://www.spotify.com>

<sup>2</sup> <http://www.apple.com/music/>

músicas foram concedidas por todos os participantes do experimento e então encaminhadas ao servidor do LIFeS. No servidor do LIFeS foi realizada a etapa de análise e processamento do áudio, para que assim pudesse ser armazenado apenas os metadados (título, autor, banda, gênero, valência, excitação, entre outros) da música.

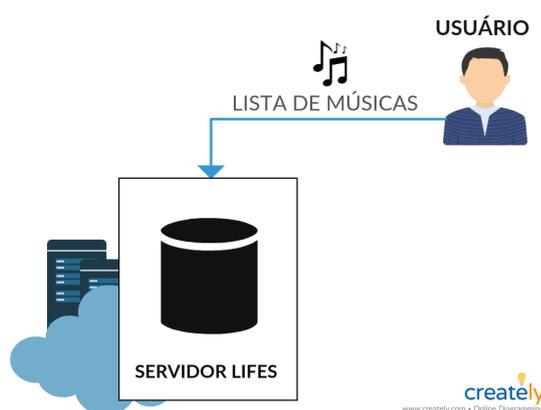
A seguir é descrito as etapas e o funcionamento real da arquitetura do m-Motion:

### 1. Músicas do Usuário

A Figura 19 representa o processo de envio de músicas para o servidor do LIFeS. Esse processo é realizado por cada usuário que pretende utilizar o m-Motion.

Portanto, para que o aplicativo m-Motion possa recomendar uma lista de músicas, bem como sua sequência, cada usuário deverá submeter um conjunto de músicas ao servidor LIFeS. Dessa forma, as músicas terão suas emoções recuperadas e armazenadas no servidor. A submissão pode ser realizada via formulário web ou enviada ao próprio pesquisador.

**Figura 19 – Análise das músicas do usuário por meio do Servidor do LIFeS.**



**Fonte:** Elaborada pelo autor.

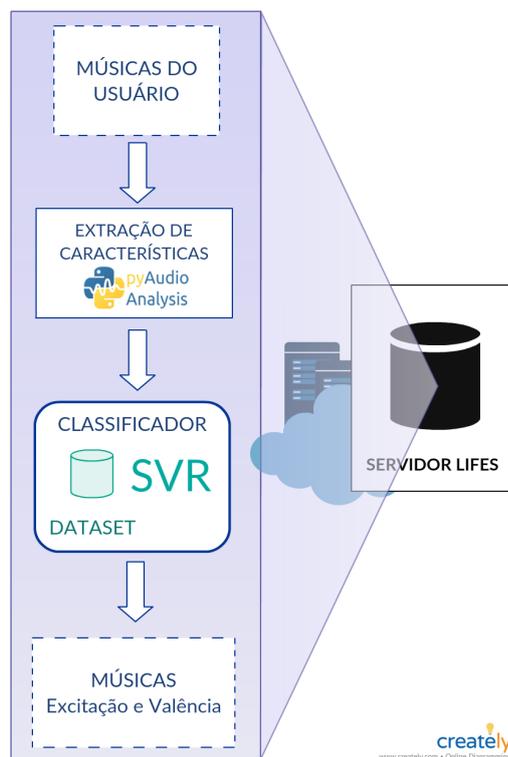
### 2. Servidor LIFeS

Quando as músicas são recebidas no servidor LIFeS, um processo é realizado para que a emoção de cada música possa ser armazenada no servidor, veja a Figura 20.

De cada música do usuário é extraído 34 recursos de áudio através da biblioteca pyAudioAnalysis<sup>3</sup>. No contexto do reconhecimento de emoções das músicas, o estado emocional não é uma classe discreta, mas sim uma medida de valor real (por exemplo, excitação e valência). Portanto, o método de análise de regressão é considerado para identificar as emoções. Logo, a biblioteca possui a capacidade de permitir treinar e extrair modelos de regressão otimizados para cada parâmetro aprendido durante a extração de características do áudio.

<sup>3</sup> pyAudioAnalysis - <https://github.com/tyiannak/pyAudioAnalysis/wiki>

**Figura 20 – Funcionamento do servidor do LIFeS.**



**Fonte:** Elaborada pelo autor.

A Tabela 3 apresenta a lista completa dos 34 recursos implementados e a sua descrição:

**Tabela 3 – Número total de recursos de curto prazo implementados pela biblioteca pyAudioAnalysis.**

ID do recurso	Nome do Recurso	Descrição
1	Taxa de cruzamento zero	A taxa de alterações de sinal do sinal durante a duração de um quadro específico.
2	Energia	A soma dos quadrados dos valores do sinal, normalizados pelo respectivo comprimento do quadro.
3	Entropia de Energia	A entropia das energias normalizadas dos subquadros. Pode ser interpretado como uma medida de mudanças abruptas.
4	Centroid espectral	O centro de gravidade do espectro.
5	Propagação Espectral	O segundo momento central do espectro.
6	Entropia Espectral	Entropia das energias espectrais normalizadas para um conjunto de sub-quadros.
7	Fluxo Espectral	A diferença quadrática entre as magnitudes normalizadas dos espectros dos dois quadros sucessivos.
8	Rolloff espectral	A frequência abaixo da qual 90% da distribuição de magnitude do espectro está concentrada.
9 a 21	MFCCs	Os Coeficientes Cepstrais de Frequência Mel formam uma representação cepstral onde as bandas de frequência não são lineares, mas distribuídas de acordo com a escala de mel.
22 a 33	Vector Chroma	Uma representação de 12 elementos da energia espectral, em que as caixas representam as 12 classes de notas temperadas iguais da música do tipo ocidental (espaçamento de semitons).
34	Desvio Chroma	O desvio padrão dos 12 coeficientes cromáticos.

**Fonte:** (GIANNAKOPOULOS, 2015)

Após extrair os 34 recursos do sinal da música do usuário, esses recursos são aplicados ao classificador SVR (Support Vector Regression), que a princípio foi treinado para prever os valores de excitação e valência para cada uma das músicas desconhecidas do usuário.

O modelo treinando para este trabalho utiliza o dataset criado por Soleymani e colaboradores (SOLEYMANI et al., 2013), com músicas obtidas da Free Music Archive (FMA<sup>4</sup>).

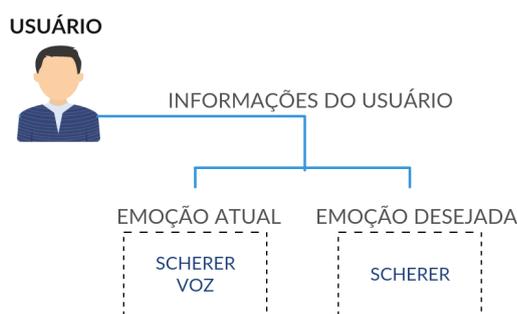
O dataset possui 1000 músicas selecionadas da FMA. Todas as músicas possuem anotações sobre suas informações, por exemplo, excitação, valência, médias com taxa de amostragem de 2Hz e desvio padrão para que possa ter uma ideia sobre a margem de erro. Além do mais cada música possui um tamanho limitado a 45 segundos.

Após prever os valores de excitação e valência da cada uma das músicas do usuário, essas informações são armazenadas no servidor do LIFeS juntamente com outros metadados (título, autor, album, etc). Apenas os metadados das músicas são armazenados no servidor e não os arquivos de áudio. Desta forma é possível evitar qualquer violação dos direitos autorais sobre o armazenamento e reprodução das músicas.

### 3. Informações do Usuário:

A Figura 21 mostra que o usuário fornece duas informações para a aplicação m-Motion. A primeira delas é a emoção atual, na qual é capturada por um componente visual do espaço semântico de Scherer e o texto da voz.

**Figura 21 – Dados de entrada dos usuários capturados pelo m-Motion.**



create.ly  
www.create.ly.com • Online Diagramming

**Fonte:** Elaborada pelo autor.

A segunda informação recuperada é a emoção desejada, e esta é capturada após ter obtido a emoção atual. A emoção desejada é obtida somente por meio do componente visual de Scherer na aplicação móvel.

### 4. API Senpy

<sup>4</sup> FMA - <http://freemusicarchive.org/>

Após gravar o áudio da voz do usuário, na qual ele informa por um discurso o seu estado emocional atual, a parte textual do áudio é extraída e enviada para a API Senpy, veja a Figura 22, em seguida a emoção que corresponde ao texto da voz é retornada.

**Figura 22 – Recuperação da emoção do usuário por meio do texto da voz.**



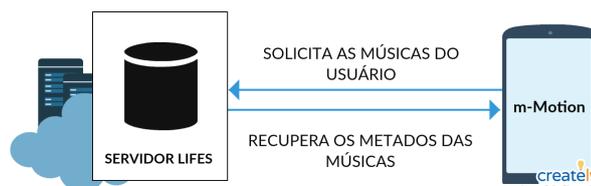
**Fonte:** Elaborada pelo autor.

A API Senpy é uma estrutura para análise de texto que permite que você crie serviços da Web de análise de sentimentos com facilidade usando dados vinculados. Existem três aplicações principais que compõem a Senpy: análise de sentimentos e emoções, perfis de usuários e reconhecimento de entidades. Com isso, torna-se possível que a aplicação m-Motion obtenha a emoção atual do usuário por meio do texto extraído da voz.

## 5. Recuperação dos metadados das músicas do usuário

Em paralelo a requisição que obtém a emoção do texto do discurso do usuário, uma requisição é realizada ao servidor do LIFeS para recuperar os metadados da lista de músicas do usuário, conforme é mostrado na Figura 23.

**Figura 23 – Recuperação dos metadados das músicas do usuário.**



**Fonte:** Elaborada pelo autor.

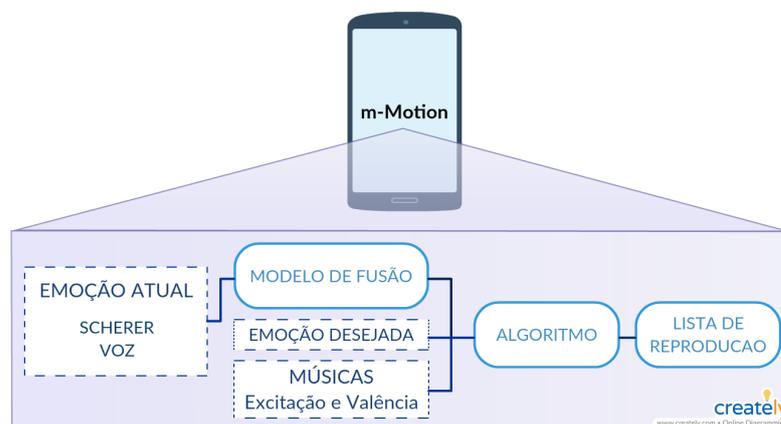
O servidor do LIFeS retorna para a aplicação m-Motion apenas os metadados das músicas e não o arquivo de áudio. Sendo assim, o usuário que interage com a aplicação recebe apenas as suas músicas com os seguintes metadados: título, autor, álbum, excitação, valência e uma url externa para reprodução.

## 6. m-Motion

Uma das principais partes da aplicação m-Motion é o algoritmo de recomendação, que é responsável por gerar a lista e a sequência das músicas do usuário. O detalhamento do comportamento do algoritmo pode ser visto na Seção 4.4. O algoritmo recebe como parâmetro três entradas: a emoção atual, a emoção desejada e uma lista de músicas

do usuário. No entanto, a emoção atual é composta de duas entradas e o algoritmo de recomendação espera uma única entrada para a emoção atual. Sendo assim, é necessário que seja feita uma fusão das duas emoções obtidas para o estado emocional atual, veja a Figura 24.

**Figura 24 – Funcionamento do aplicativo m-Motion**



**Fonte:** Elaborada pelo autor.

O **modelo de fusão** recebe como entrada a emoção recuperada pelo componente de Scherer e a emoção do texto da voz recuperada pela API Senpy. Com isso, a fusão é realizada por regras de prioridade. A fusão de regras de prioridade define cada entrada com um determinado peso de prioridade e tende a retornar uma saída próxima as entradas de maiores pesos. Na aplicação m-Motion, o componente de Scherer foi definido como a entrada de maior peso. Isto se justifica, pois o componente de Scherer é mais preciso do que a emoção obtida pelo texto da voz. Com o componente de Scherer é possível retornar instantaneamente os pontos de excitação e valência, enquanto para obter emoção do texto da voz é necessário realizar uma requisição a outro servidor. Além do mais, foi possível perceber que a API Senpy possui um vocabulário de emoções limitado. Portanto, a fusão das duas entradas da emoção atual é realizada da seguinte maneira: primeiro é calculado  $M$ , sendo que  $M$  é o ponto médio entre a distância da coordenada da emoção atual com a desejada, em seguida é calculado  $MF$  que é outro ponto médio, no entanto entre a distância da emoção atual com o ponto médio anteriormente calculado  $M$ . Com isso  $MF$  representa a fusão final da emoção atual.

Dessa forma, com a fusão da emoção atual, a emoção desejada e a lista de músicas do usuário sob controle da aplicação m-Motion, o algoritmo tem as entradas necessárias para processar e retornar uma lista de reprodução para o usuário.

### 4.3 Algoritmo

O procedimento para selecionar as músicas que pretendem levar o usuário ao estado emocional desejado é descrito no Algoritmo 1.

**Algorithm 1** Seleção de Música

**Entrada:** Lista de músicas (playlist), Emoção Atual (CurrentEmo) e Emoção desejada (desiredEmo)

**Saída:** Lista de músicas selecionadas (newPlaylist)

**início**

```

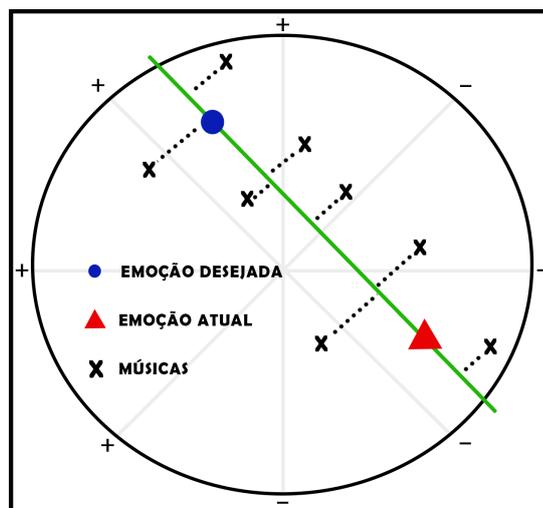
// Calcula da distância das músicas até a reta
for  $i = 0$  to  $\text{size}(\text{playlist})$  do
  |  $v\text{Distance}[i] = d\text{PointLine}(\text{currentEmo}, \text{desiredEmo}, \text{playlist}[i])$ 
end
// Acessa o vetor de distância de músicas
for  $i = 0$  to  $\text{size}(v\text{Distance})$  do
  // Detecta músicas com distância menor que 0,3 da reta
  if  $v\text{Distance}[i] < 0.3$  then
    // Limita a busca de músicas ao longo da reta
    if  $\text{limit}(\text{currentEmo}, \text{desiredEmo})$  then
      |  $\text{newPlaylist}[i] = \text{playlist}[i]$ 
    end
  end
end
end
return newPlaylist

```

**fim**

O algoritmo recebe como entrada três parâmetros, sendo *CurrentEmo* e *desiredEmo* coordenadas que referenciam a excitação e valência da emoção atual e desejada do usuário. O terceiro parâmetro, *playlist*, é uma matriz que contém todas as coordenadas de excitação e valência das músicas do usuário que já estão classificadas no espaço emocional semântico com valores reais entre -1 e 1.

**Figura 25 –** Comportamento da função *dPointLine*



**Fonte:** Elaborada pelo autor.

A Figura 25 apresenta o comportamento da função *dPointLine*, na qual realiza duas operações, a primeira calcula a equação da reta entre a emoção atual e emoção desejada, que é

representado pela reta em verde e a segunda operação calcula a distância de uma determinada música com a reta, que é representado pela linha pontilhada. O primeiro laço de repetição realiza o cálculo de distância de todas as músicas do usuário com a reta que foi criada, e então armazena todas as distâncias no vetor  $vDistance$ .

O segundo laço de repetição percorre o vetor  $vDistance$  e detecta músicas que possuem uma distância menor do que 0.3 da reta. Com isso, o novo conjunto de músicas é constituído apenas de músicas com distância inferior a 0.3 ao longo da reta. A distância 0.3 é um parâmetro que foi definido considerando o alcance do espaço emocional semântico que é limitado com valores reais entre -1 e 1. Com esse valor, é possível assegurar que o percurso da reta não envolverá todo espaço semântico emocional semântico, bem como não incluirá uma grande quantidade de músicas próximas a emoção atual.

Com o novo conjunto constituído de músicas com distância inferior a 0.3 ao longo da reta, é necessário realizar um novo limite para esse conjunto de forma que satisfaça as seguintes condições:

1. Se o estado emocional desejado não for próximo ao estado atual, então deve-se evitar que músicas próximas ao estado emocional atual sejam retornadas, visto que as músicas próximas ao estado atual tendem a manter o usuário no mesmo estado emocional.
2. Evitar que as músicas superiores ou fora do raio da emoção desejada sejam retornadas, pois como a reta que corta a emoção atual e desejada tende ao infinito, isso faz com que músicas fora do percurso entre a emoção atual e desejada também sejam retornadas, pois estabelecem a distância de 0.3 da reta.

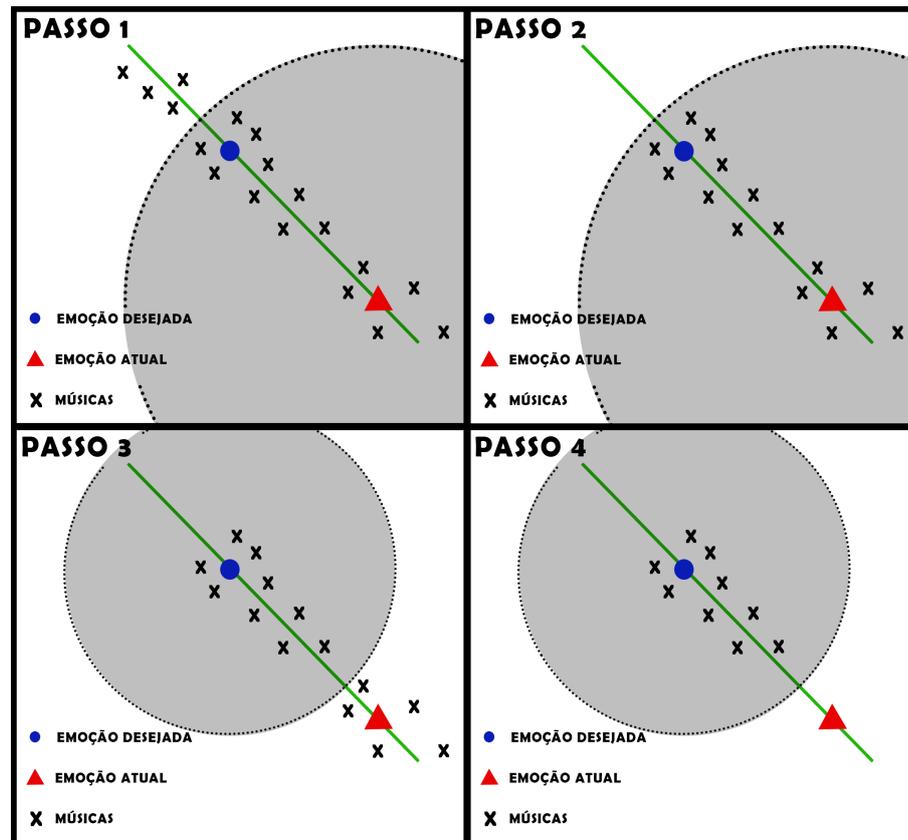
Essas condições são resolvidas com a função *limit*, veja o comportamento da função na Figura 26.

A função é capaz de criar um limite para a reta, partindo da emoção atual até a emoção desejada. No passo 1, a função cria um círculo em torno da emoção atual com raio definido como R1. R1 é definido como:

$$R1 = (Distancia_{Atual e Desejada} + 0.3)$$

Onde, R1 é a soma do valor da distância entre a coordenada da emoção atual e desejada com o parâmetro 0.3. Em seguida, no passo 2, todas as músicas fora do raio R1 são removidas. Com isso, as músicas superiores ou fora do raio da emoção desejada não são retornadas e a condição 2 é satisfeita.

No passo 3, a função cria um círculo em torno da emoção desejada com raio definido por R2. R2 é definido como:

Figura 26 – Comportamento da função *limit*

Fonte: Elaborada pelo autor.

$$R2 = (Distancia_{Atual e Desejada} - 0.3)$$

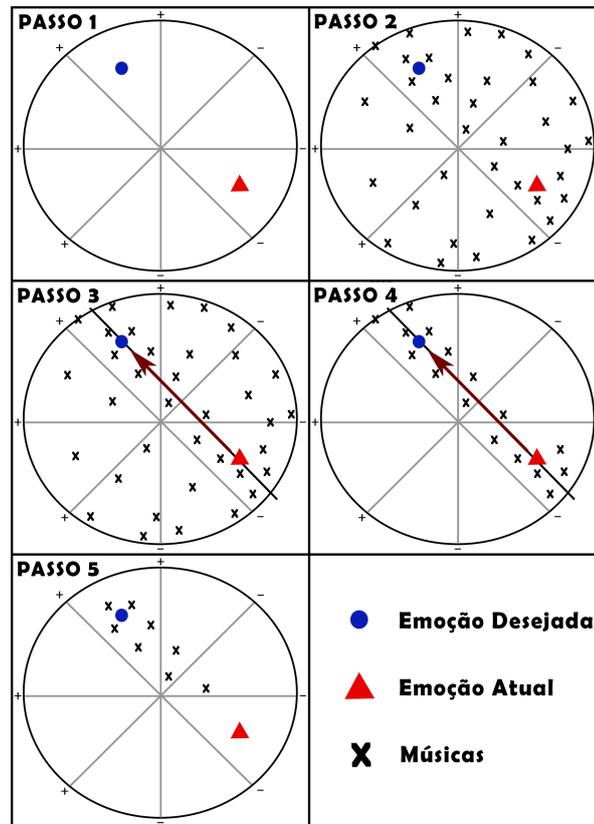
Onde, R1 é a subtração do parâmetro 0.3 do valor da distância entre a coordenada da emoção atual e desejada. Posteriormente, no passo 4, todas as músicas fora do raio R2 são removidas. Desta maneira, as músicas próximas ao estado emocional atual são ignoradas e a condição 1 é satisfeita.

Por fim, o algoritmo retorna para a aplicação principal a lista de músicas que serão capazes de levar o usuário para o estado emocional desejado. Caso não seja encontrada nenhuma música entre o percurso da emoção atual com a desejada, o algoritmo retorna o conjunto de música classificada em ordem crescente, tendo como parâmetro a distância da música com a emoção desejada.

A partir da Figura 27 é possível ver a ilustração de um exemplo do comportamento do algoritmo para um determinado caso.

- No passo 1 e 2 o algoritmo obtém a emoção atual, desejada e a lista de músicas do usuário com suas emoções já classificadas em termos de excitação e valência.
- O passo 3 cria a equação da reta que passa pelos pontos da emoção atual e desejada.

Figura 27 – Etapas para o algoritmo para selecionar músicas.



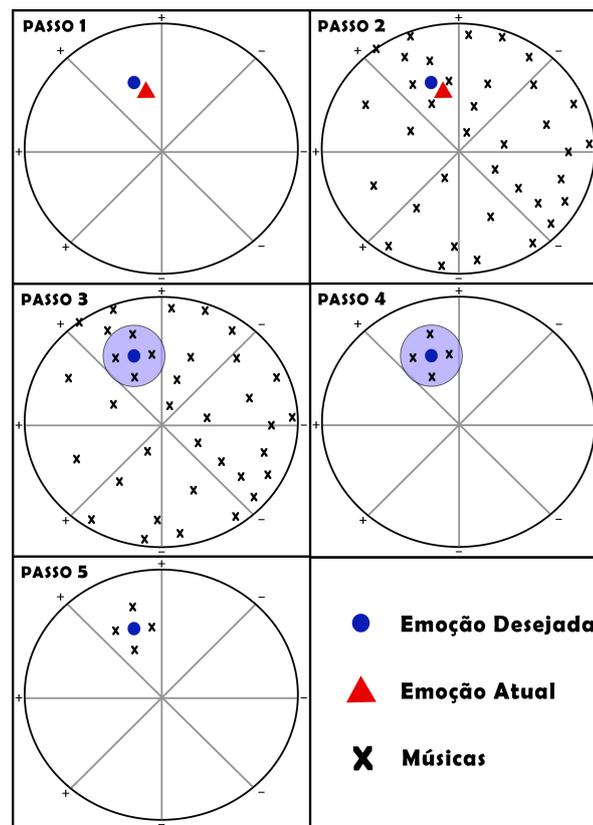
Fonte: Elaborada pelo autor.

- Em seguida, no passo 4, as músicas que possuem distância maior ou igual a 0.3 são ignoradas. O conjunto de músicas remanescentes está localizado ao longo de toda a reta.
- Por fim, no passo 5 é criado o limite na reta. Músicas próximas a emoção atual, bem como superiores à emoção desejada são ignoradas. Com isso, o algoritmo retorna ao usuário uma lista de músicas tidas como capaz de levá-lo para o estado emocional desejado.

Um segundo exemplo do comportamento do algoritmo é apresentado na Figura 28. Nesse exemplo a emoção desejada encontra-se próxima da emoção desejada.

- No passo 1 e 2 o algoritmo obtém a emoção atual, desejada e a lista de músicas do usuário com suas emoções já classificadas em excitação e valência.
- O passo 3 o algoritmo verifica que a emoção desejada está próxima da emoção atual, portanto, não cria a equação da reta entre a emoção atual e desejada, mas sim um círculo com raio de 0.3 da emoção desejada.
- Em seguida, no passo 4, as músicas que possuem distância maior ou igual a 0.3 da emoção desejada são ignoradas.

**Figura 28 – Etapas para o algoritmo para selecionar músicas.**



Fonte: Elaborada pelo autor.

- Por fim, no passo 5 o algoritmo retorna ao usuário uma lista de músicas remanescentes de todo o processo que são tidas como capaz de levá-lo para o estado emocional desejado.

## 4.4 Reprodutor mobile

A conclusão deste trabalho de mestrado gerou uma ferramenta para dispositivos móveis que busca ajudar o usuário alcançar um estado emocional desejado, considerando suas músicas e seu estado atual. A aplicação foi desenvolvida utilizando uma tecnologia híbrida para as duas plataformas Android e iOS. As emoções foram recuperadas utilizando 3 artefatos: texto da voz, componente visual de Scherer e componente visual de SAM (Self Assessment Manikin). A seguir, nas Subseções seguintes são apresentadas as interfaces da aplicação, os componentes criados para captura das emoções, bem como a descrição do seu comportamento.

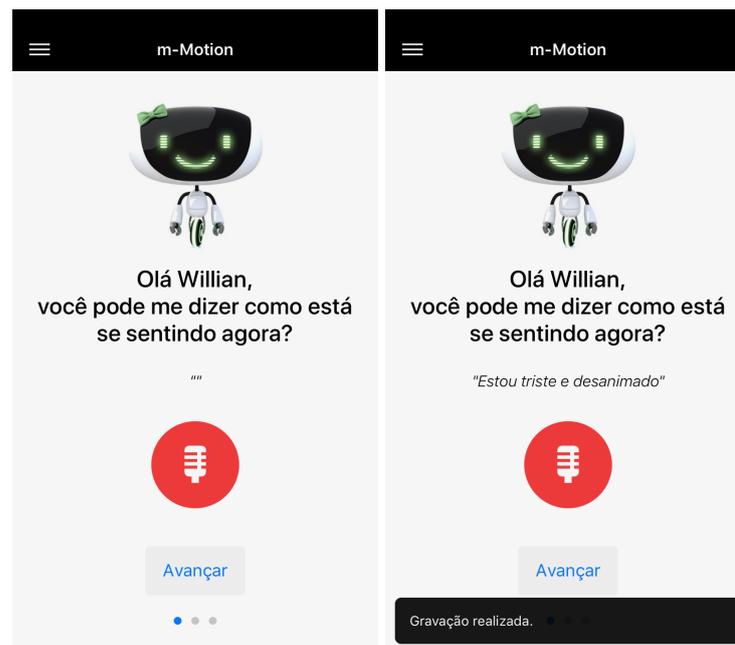
### 4.4.1 Interface de usuário para recuperação da emoção atual e desejada do usuário

A emoção atual do usuário é obtida por dois artefatos: pelo texto da voz e por um componente visual inspirado no espaço semântico de Scherer. A Figura 29 apresenta a interface

de captura do texto da voz da aplicação m-Motion. Nessa tela, o usuário inicia uma gravação de voz, no qual discursa sobre como está se sentindo no momento da interação.

Após a concluir a gravação, a aplicação converte o discurso do usuário em texto e exibe na tela. Em paralelo, a aplicação realiza duas operações internas: envia o texto do discurso para a API Senpy e recupera a emoção do texto, bem como solicita os metadados das músicas do usuário que realiza a interação e armazena para processá-los posteriormente.

**Figura 29 – Interface visual para captura do texto da voz do m-Motion.**



**Fonte:** Elaborada pelo autor.

Para obter as emoções do usuário de forma subjetiva e via dispositivo móvel, um componente visual inspirado no espaço semântico de Scherer foi desenvolvido. Através do componente o usuário realiza um toque em uma determinada área do espaço e os valores de excitação e valência correspondentes são recuperados. O componente possui 4 ícones ilustrativos de expressões faciais, assim como a divisão do espaço em 4 grupos com cores diferentes, de maneira que possa ajudar o usuário localizar visualmente uma determinada emoção no espaço do componente.

**Figura 30 – Componente visual Scherer. Adaptado do modelo do espaço semântico de Scherer para captura da emoção atual e desejada via dispositivo móvel.**

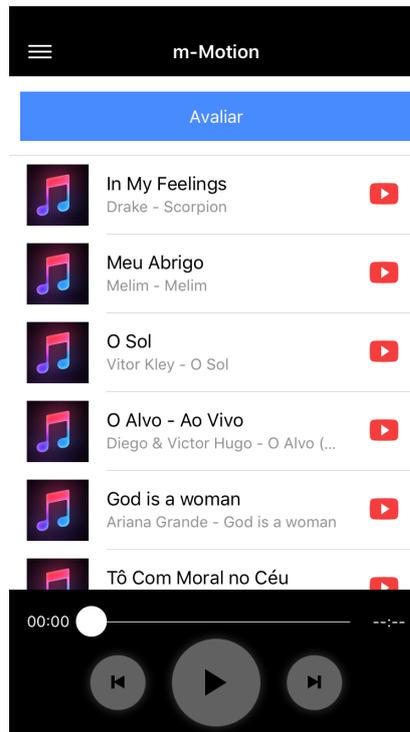


**Fonte:** Elaborada pelo autor.

A Figura 30 mostra o componente visual utilizado para capturar a emoção atual e desejada do usuário. Na primeira tela, o usuário realiza uma marcação no componente que referencia o seu estado emocional atual, representado pelo círculo em azul. A emoção desejada é obtida apenas pela componente visual de Scherer e sua representação pode ser vista na segunda tela.

#### 4.4.2 Interface de usuário da lista de reprodução

Após capturar a emoção atual e desejada do usuário por meio dos componentes, a aplicação fica preparada para gerar a lista de reprodução para o usuário, visto que nas etapas anteriores os metadados das músicas já foram recuperados. Portanto, ao ser guiado para a tela da lista de reprodução, o algoritmo de recomendação recebe as 3 entradas necessárias para o seu funcionamento e realiza todo o processo prescrito para retornar a lista de reprodução final. A Figura 31 apresenta a interface da tela do reprodutor móvel com a lista recomendada.

**Figura 31 – Interface da lista de reprodução do m-Motion.**

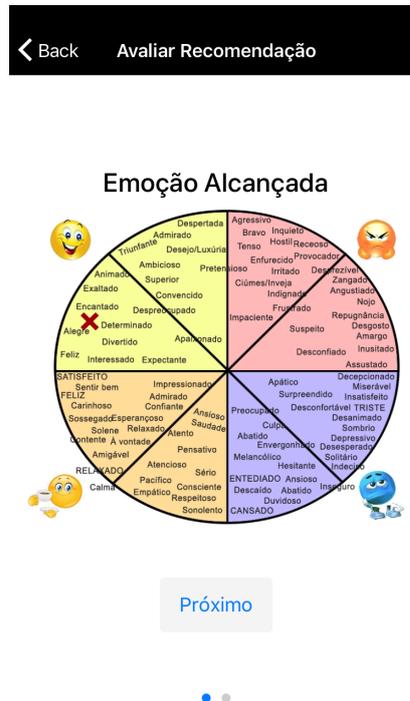
**Fonte:** Elaborada pelo autor.

A lista de reprodução recomendada exibe o título da música, autor, álbum e link externo para reprodução. Buscando evitar a violação dos direitos autorais sob o armazenamento e reprodução das músicas sugeridas pelos usuários, este trabalho optou por recomendar a reprodução das músicas em plataformas externa e que garante os direitos autorais sobre as músicas. Desta maneira, o usuário recebe a lista de recomendação e fica capaz de reproduzi-las em sua plataforma de preferência.

#### 4.4.3 Interface de usuário de avaliação da emoção alcançada

Após a reprodução completa da lista de músicas recomendada, o usuário tem a possibilidade de avaliar a recomendação sugerida. A avaliação é realizada por meio dos componentes de Scherer e SAM, para que assim possa verificar a emoção alcançada pelo usuário após ouvir as músicas. As Figuras 32 e 33 apresentam a interface das telas de avaliação da aplicação m-Motion.

**Figura 32 – Componente visual Scherer. Adaptado do modelo do espaço semântico de Scherer para captura da emoção alcançada do usuário via dispositivo móvel.**



**Fonte:** Elaborada pelo autor.

Para verificar o estado emocional do usuário após a reprodução das músicas e obter o valor de excitação e valência da emoção correspondente, o componente de Scherer foi utilizado, visto que ele permite que o usuário visualize uma emoção no espaço semântico e realize uma marcação.

Além do componente de Scherer, outro artefato é utilizado em paralelo para mensurar a emoção alcançada pelo usuário. O componente de SAM foi desenvolvido para obter as emoções do usuário de forma subjetiva e via dispositivo móvel, um componente visual inspirado na técnica de auto-relato Self Assessment Manikin (SAM). O componente possui representações imagéticas (os Manikins) para duas dimensões: excitação e valência. As duas dimensões possuem um componente deslizante que é capaz de ajustar e selecionar um Manikin que melhor representa o estado emocional do usuário. Desta maneira é possível obter um valor de excitação e valência que representa a emoção do estado alcançado pelo usuário.

Após o usuário informar a sua emoção alcançada pelos dois componentes, as informações são enviadas ao servidor do LIFeS para analisar o resultado das recomendações realizadas pelo m-Motion.

**Figura 33 – Componente visual SAM. Adaptado da técnica self-assessment manikin (SAM) para captura da emoção alcançada do usuário via dispositivo móvel.**

← Back Avaliar Recomendação

### Emoção Alcançada

EXCITAÇÃO: 50%

-100% 100%

VALÊNCIA: 75%

-100% 100%

Voltar Enviar

• •

**Fonte:** Elaborada pelo autor.

# Capítulo 5

## AVALIAÇÃO

---

---

### 5.1 Considerações Iniciais

O objetivo principal deste trabalho é ajudar o usuário alcançar um estado emocional desejado por meio das emoções musicais, ou seja, verificando o estado emocional atual do usuário, em excitação e valência, o algoritmo proposto deve retornar suas músicas preferidas que se aproximam do estado emocional desejado. Para isso, foi realizado um experimento com a finalidade de avaliar a emoção alcançada pelo usuário após ouvir uma seleção de músicas sugeridas. O experimento foi realizado com 3 grupos distintos de participantes, sendo dois grupos com 20 participantes e um grupo de 8 participantes, dos quais possuíam idades entre 18 e 50 anos. Cada grupo participou do experimento em períodos diferentes entre si. Além do mais, cada participante assinou um termo de consentimento de maneira que pudessem tomar suas decisões de forma justa e sem constrangimentos. De cada participante dos grupos formados com 20 integrantes foi coletado o estado emocional atual, desejado e 20 sugestões de músicas. O grupo com 8 participantes, a emoção atual e a desejada foi coletada pela própria aplicação m-Motion durante o experimento em ambiente não controlado. Além disto, a lista dos participantes eram compostas por 20 músicas de sua preferência.

### 5.2 Planejamento dos Experimentos

Para o planejamento dos experimentos exploratórios, foram definidos os objetivos, a hipótese, a metodologia, o material e os dispositivos utilizados para verificação da hipótese.

*Hipótese:* Um reproduzidor de música que recupera a emoção das músicas do usuário, além de seus sentimentos subjetivos para fazer uma recomendação, que é baseada em cálculos matemáticos, por exemplo, a equação da reta entre um ponto da emoção atual e desejada, junto com a distância euclidiana pode ajudar o usuário atingir um estado emocional desejado.

*Objetivo:* coletar e analisar respostas emocionais de usuários antes, durante e após a reprodução de uma lista de música.

*Metodologia Aplicada:* com o intuito de analisar e coletar respostas emocionais do usuário durante a reprodução de uma lista de música, três grupos de usuários foram convidados para participar de um experimento que envolvia o uso de um computador ou um dispositivo móvel para reprodução das músicas. O primeiro grupo, os participantes tinham como objetivo ouvir uma sequência de música retornada pelo algoritmo proposto, de maneira que essa sequência pudesse levá-los para o estado emocional desejado informado. O segundo grupo, os participantes tinham como objetivo sugerir de forma manual uma sequência de músicas que acreditassem levá-los para um estado emocional desejado. O objetivo da atividade do terceiro grupo consistia na reprodução de uma determinada lista de reprodução retornada pela aplicação m-Motion, portanto, essa tarefa foi realizada utilizando dispositivos móveis (celulares).

*Material de Apoio:* para a condução dos experimentos foram preparados: um Termo de Consentimento Livre e Esclarecido, um Termo de Autorização de Captação e Exibição de Imagem, Som e Nome, um artefato com a representação do Espaço Semântico de Scherer e o instrumento de avaliação (auto-relato) chamado Self Assessment Manikin (SAM) . O Termo de Consentimento Livre e Esclarecido elucidada, aos participantes, o objetivo da pesquisa, a voluntariedade na participação e o caráter científico da mesma (ver Apêndice A). O Termo de Autorização de Captação e Exibição de Imagem, Som e Nome esclarece aos participantes que há captura de imagens e voz e que os dados coletados são utilizados para fins científicos (ver Apêndice B). O artefato do Espaço Semântico de Scherer permite que os usuários realizem uma marcação em uma determinada região do espaço próximo a uma determinada emoção que se encontra ou deseja alcançar (ver Apêndice C). O instrumento de SAM permite que o usuário marque um valor dentre a escala de 1 a 9 para a excitação e valência, de maneira que essa escala represente a emoção provocada após a reprodução das músicas (ver Apêndice D).

### 5.2.1 Critérios de Inclusão de Usuários nos Experimentos

Os experimentos utilizando a recomendação manual pelos próprios usuários e por meio do algoritmo proposto, os participantes foram convidados por e-mail enviado aos departamentos da universidade e amplamente divulgado para a comunidade externa à universidade, além de divulgação pelas redes sociais. A mensagem conteve um breve resumo da pesquisa: objetivos, local onde será realizado o experimento, tarefa a ser realizada, tempo estimado, esclarecimentos sobre a participação e contato do pesquisador.

O experimento utilizando a recomendação pela aplicação m-Motion, a escolha dos participantes foram feitas de forma não aleatória. A escolha foi realizada por conveniência, considerando as características particulares do grupo em estudo e conhecimento do que está investigando (OLIVEIRA, 2001).

O recrutamento ocorreu durante dois meses e atingiu 48 participações, pois, para Nielsen (2006) estudos quantitativos visando estatísticas de usabilidades são necessários pelo menos 20 usuários para obter números estatisticamente significativos. Os critérios de inclusão dos

sujeitos na pesquisa foram: maiores de 18 anos; brasileiros; e alfabetizados. Como se tratava de uma atividade que exigia a visualização de imagens e textos, adotou-se como exclusão a não participação de pessoas cegas ou com baixa visão. Os dados dos participantes que por livre espontânea vontade não concluíram os experimentos não foram considerados na análise.

### **5.2.2 Riscos em Participar dos Experimentos**

Existem os seguintes riscos para os participantes do trabalho: estresse, tristeza, cansaço e tédio, seja pelos conteúdos, tempo de realização das tarefas e/ou resposta de questionários. O sujeito de pesquisa pode, ainda, se frustrar pela recomendação de músicas sugeridas por algum dos métodos, como também se sentir desconfortável por estar sendo avaliado.

Será feito o possível para minimizar os impactos listados previamente. O participante poderá estar em um estado emocional de baixa excitação com a intenção de ouvir músicas e que animam, no entanto, caso ocorra a recomendação de músicas desagradáveis aos participantes, estas serão excluídas do experimento. Ainda assim, caso ocorram, o participante pode se recusar a responder aos questionários, como também para a escuta das músicas ou até mesmo interromper a sua participação a qualquer momento.

### **5.2.3 Benefícios em Participar dos Experimentos**

Os benefícios relacionados à participação dos sujeitos de pesquisa estão em contribuir com o trabalho, assim como para o estado da arte na recomendação de músicas do usuário com base no estado emocional atual e desejado do usuário. Além disso, a partir dos experimentos foi possível aprimorar a abordagem para estimar o estado emocional de indivíduos na interação com sistemas computacionais. Os sujeitos da pesquisa puderam contribuir também com a inclusão social e digital, haja vista que o estado emocional pode ser usado como artifício para melhorar a forma como as soluções computacionais são apresentadas à população.

## **5.3 Recomendação pelo Algoritmo e Usuário**

Os experimentos com recomendação de músicas pelo algoritmo e pelo próprio usuário tiveram 20 usuários cada grupo, sendo que ambos foram realizados em períodos diferentes e em um ambiente controlado. Os dois experimentos tinham como objetivo avaliar a emoção alcançada pelos participantes após a reprodução de um conjunto de músicas sugeridas. No entanto, a diferença entre os dois experimentos está na forma como as músicas foram sugeridas. O primeiro experimento a recomendação foi realizada pelo algoritmo. Dentre as 20 músicas sugeridas previamente de cada participante, o algoritmo realizou uma seleção de maneira que essa seleção pudesse levar o usuário para um estado emocional desejado. O segundo experimento a recomendação foi realizada pelo próprio participante. Dentre um conjunto de 20 músicas do

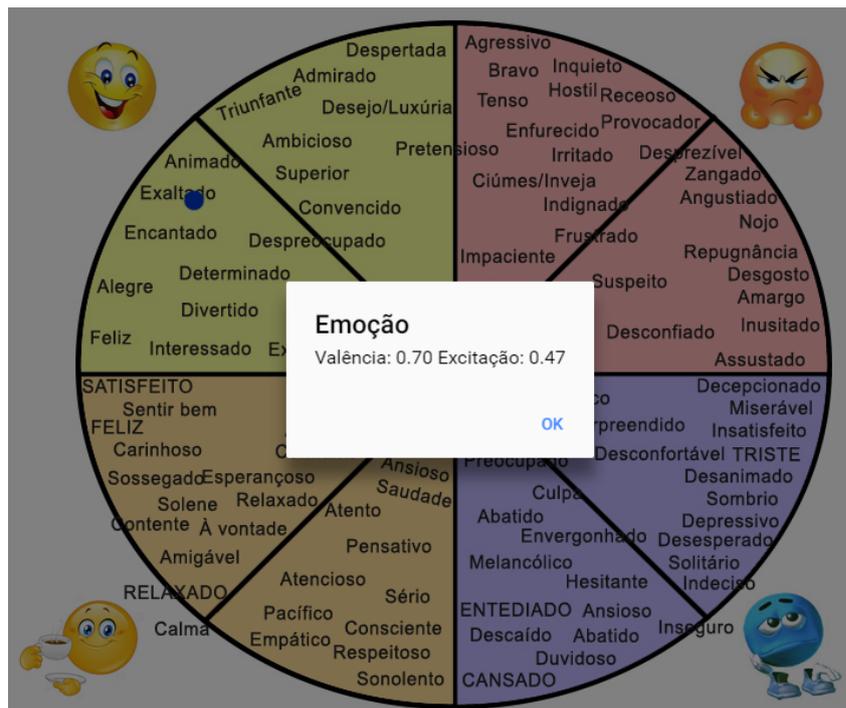
participante, o mesmo realizava uma seleção a partir das 20 músicas, de forma que sua seleção pudesse levá-lo para um estado emocional desejado.

As subseções seguintes apresentam etapas análogas aos dois experimentos: recomendação pelo algoritmo e recomendação pelo próprio participante.

### 5.3.1 Emoção Atual e Desejada do Participante

No início de cada experimento, tanto para o de recomendação pelo algoritmo quanto o de recomendação pelo próprio participante, cada participante recebeu um formulário com a apresentação do espaço semântico de Scherer (ver Apêndice C). A atividade consistia em marcar um ponto no espaço que representava sua emoção atual como também um segundo ponto em que representava a emoção que se desejava alcançar. As mesmas marcações realizadas pelos participantes foram aplicadas em uma interface web com o círculo do espaço semântico de Scherer, veja a Figura 34.

**Figura 34 – Interface web do espaço semântico.**



**Fonte:** Elaborada pelo autor.

O espaço semântico da interface possui um intervalo entre -1 e 1 de tal maneira que com um clique em uma determinada região tornava-se possível recuperar os valores de excitação e valência correspondentes as marcações feitas no formulário.

### 5.3.2 Reconhecimento Emocional da Música

O experimento em que os participantes realizavam a própria sugestão de músicas para alcançar um estado emocional desejado, não foi necessário a etapa de reconhecimento emocional

da música, visto que a seleção das músicas eram feitas manualmente pelo próprio participante dentre um conjunto de músicas conhecidas e de sua preferência.

Em contrapartida, o experimento de recomendação pelo algoritmo a etapa de reconhecimento emocional da música é de alta importância, pois, o algoritmo de recomendação necessita da emoção da música em excitação e valência para realizar as sugestões.

### 5.3.2.1 Lista de Reprodução dos Participantes

Antes da execução dos experimentos, cada um dos 20 participantes que atuaram em ambos experimentos, o de recomendação pelo algoritmo e pelo próprio participante, previamente sugeriram 20 músicas de sua preferência nas quais eram comumente ouvida no seu dia a dia. As músicas tiveram a emoção recuperada em valores de excitação e valência apenas para o experimento de recomendação pelo algoritmo.

### 5.3.2.2 Extração de características

Para recuperar a emoção das músicas dos participantes que atuaram no experimento de recomendação pelo algoritmo, a biblioteca pyAudioAnalysis foi escolhida, isto porque é uma biblioteca open-source escrita em Python que oferece uma ampla gama de procedimentos para análise de áudio, incluindo: extração de recursos, classificação de sinais de áudio, segmentação supervisionada e não supervisionada e visualização por conteúdo (GIANNAKOPOULOS, 2015).

### 5.3.2.3 Dataset

De modo a de treinar e usar modelos supervisionados que classificam e estimam as variáveis de regressão para um arquivo de áudio desconhecido, esse experimento utilizou o dataset criado por Soleymani e colaboradores (SOLEYMANI et al., 2013), com músicas obtidas da Free Music Archive (FMA<sup>1</sup>).

A partir desse dataset foi possível treinar e extrair os modelos de regressão que estimam os valores de excitação e valência para uma música desconhecida.

## 5.3.3 Avaliação da emoção alcançada

Durante ambos experimentos, cada participante informou a sua emoção atual e desejada. No experimento da recomendação pelo algoritmo, a emoção atual e desejada, com as respectivas listas de músicas fornecidas de cada participante, foram aplicadas ao algoritmo de recomendação, que retornou para cada participante um possível conjunto de músicas que o levasse a alcançar o estado emocional desejado. Já no experimento de recomendação manual, os próprios participantes indicaram um novo subconjunto de músicas a partir das 20 músicas informadas anteriormente.

<sup>1</sup> FMA - <http://freemusicarchive.org/>

Subconjunto que para o participante a seleção era capaz de levá-lo para um estado emocional desejado.

Assim que cada participante possuía a lista de reprodução para apoiá-lo alcançar o estado emocional desejado, seja no experimento de recomendação pelo algoritmo ou manual, cada participante se posicionou em frente a uma webcam e realizou a reprodução das músicas. Durante a reprodução das músicas, as expressões faciais dos participantes foram capturadas utilizando o kit de ferramentas de reconhecimento de expressões faciais da Affectiva (MCDUFF et al., 2016), com o intuito de avaliar a variação da emoção dos participantes durante a reprodução das músicas.

Ao fim da reprodução das músicas, foi aplicado a técnica de auto-relato Self Assessment Manikin (SAM) para avaliar o estado emocional de cada participante e o espaço emocional semântico de Scherer como um artefato de registro.

No contexto desta pesquisa, o SAM foi escolhido porque esse instrumento de auto-avaliação pode avaliar os domínios diretamente correspondentes ao espaço emocional semântico de Scherer. O SAM é composto por três conjuntos de pictogramas que representam os domínios de Valência (satisfeito - insatisfeito), Excitação (motivado - relaxado) e Sentimento de controle (completamente no controle - fora de controle) (BRADLEY; LANG, 1994). Cada domínio é representado por uma escala de valores que varia de 1-9. Os participantes foram instruídos a escolher um dos cinco pictogramas ou intermediário entre os dois, na qual melhor representa o seu estado emocional.

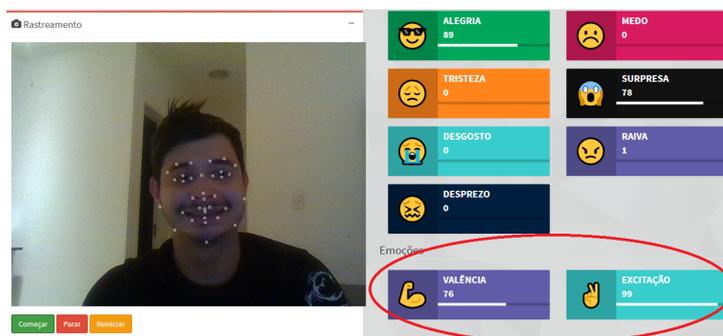
Além do SAM, foi utilizado o espaço emocional semântico de Scherer, onde cada participante realizava uma marcação em qual quadrante do espaço emocional se encontrava.

### 5.3.4 Resultados da Recomendação pelo Algoritmo

O algoritmo proposto retornou para cada um dos 20 participantes uma lista de músicas com o intuito de levar o usuário para o estado emocional desejado. As listas de reprodução não tinham uma quantidade de arquivos de áudio limitado, isso porque as músicas eram retornadas de acordo com o ponto de partida da emoção atual com a desejada.

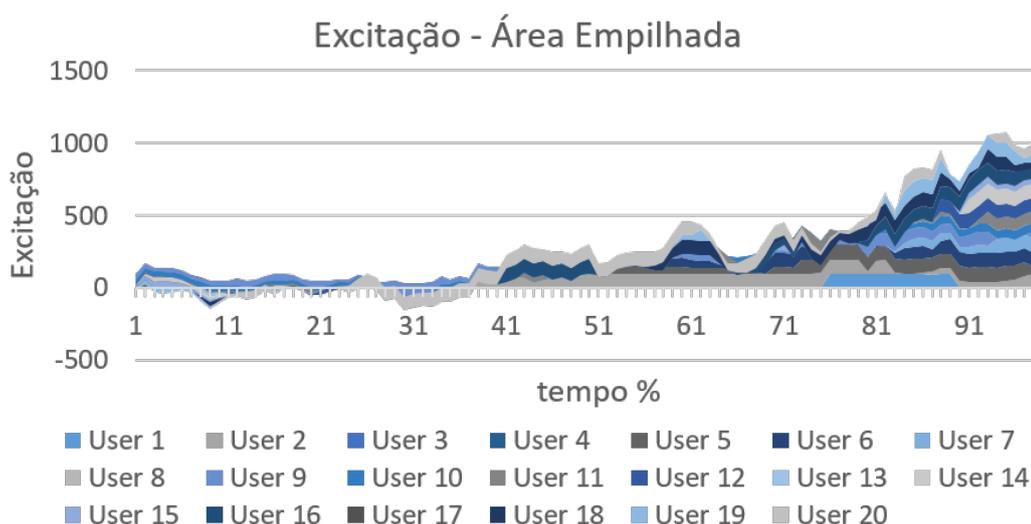
Primeiro foi avaliado as expressões faciais dos participantes utilizando o kit de ferramentas de reconhecimento de expressões faciais em tempo real da Affectiva. Com o kit da Affectiva foi possível criar uma interface web capaz de analisar as expressões faciais capturadas pela Webcam e armazenar os resultados em uma base de dados, veja a Figura 35. A Figura 35 apresenta a interface web utilizada para extrair as emoções dos participantes. O kit da Affectiva retorna os valores de excitação e valência com valores entre 0 e 100 e, portanto, uma normalização foi realizada para que os valores estivessem em um intervalo entre -1 e 1. As Figuras 36 e 37 mostram o gráfico da área empilhada no decorrer do tempo da reprodução das músicas para excitação e valência de cada participante.

**Figura 35 – Interface web para captura de emoção por meio das expressões faciais.**



**Fonte:** Elaborada pelo autor.

**Figura 36 – Total acumulativo da excitação ao longo da reprodução das listas sugeridas para cada participante.**



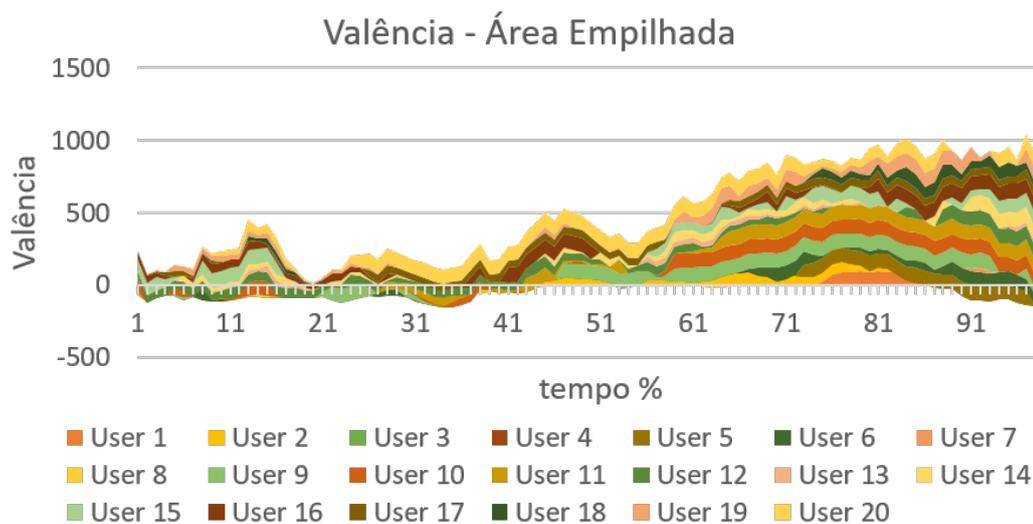
**Fonte:** Elaborada pelo autor.

Ambos os gráficos de Excitação e Valência das Figuras 36 e 37 respectivamente, mostram que a partir da reprodução de 40% das listas de músicas, os participantes tiveram a excitação e valência aumentadas gradativamente. Isso nos mostra que as músicas sugeridas pelo algoritmo proposto influenciaram os participantes em variar o seu estado emocional.

Garrido (2014) realizou uma revisão sistemática da literatura com finalidade de comparar vários métodos utilizados para medir a emoção e o humor na literatura musical e avaliar sua eficácia relativa. Uma das discussões apresentadas no trabalho é que embora uma música muito curta possa ser proveitosa para perceber as emoções, um estímulo um pouco maior pode ser necessário para induzir uma resposta emocional a um ouvinte.

A revisão apresenta alguns trabalhos que relatam que, em média, os participantes precisavam de cerca de 30 a 60 segundos para relatar uma resposta emocional à música. Em outros, uma música inteira pode não ser suficientemente para induzir qualquer alteração de emoção do ouvinte. E que para uma verdadeira indução emocional tenha sido criada, o efeito deve durar de

**Figura 37 – Total acumulativo da valência ao longo da reprodução das listas sugeridas para cada participante.**



**Fonte:** Elaborada pelo autor.

5 a 40 minutos.

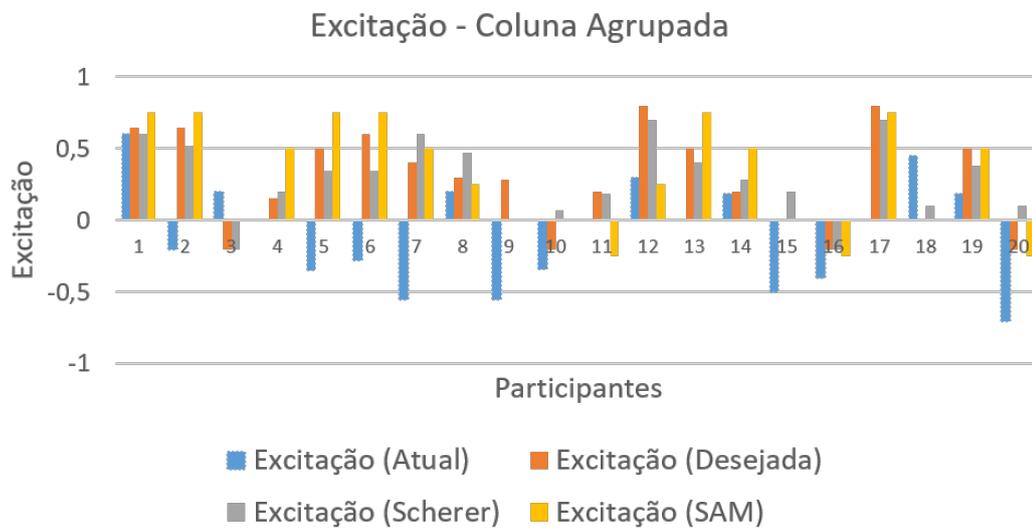
Portanto, com a revisão de Garrido (2014) é possível reforçar os resultados apresentados nas Figuras 36 e 37, onde os participantes mostraram um aumento gradualmente na excitação e valência durante a reprodução das músicas retornadas pelo algoritmo proposto. Ressalta-se que também, que os gráficos das Figuras 36 e 37 apresentam uma normalização do tamanho das listas de reprodução. As listas sugeridas não possuíam uma quantidade delimitada ou fixa de músicas, isso porque a lista era criada de acordo com as emoções das músicas do participante, bem como a emoção atual e desejada.

Após avaliar as emoções dos participantes durante a reprodução das músicas e verificar se produziam uma variação no estado emocional, verificou-se que após a escuta das músicas os participantes conseguiram se aproximar do estado emocional desejado. As Figuras 38 e 39 mostram os gráficos de excitação e valência para os estados atuais, desejados e alcançados dos participantes.

Ambos os gráficos, de excitação e valência das Figuras 38 e 39 respectivamente, possuem 4 colunas para cada participante que referenciam os estados emocionais dos participantes, sendo que, as colunas em azul representam a emoção atual antes da reprodução das músicas e as colunas em vermelho representam a emoção desejada dos participantes. As demais colunas em cinza e amarelo representam as emoções alcançadas pelos participantes após a reprodução das músicas.

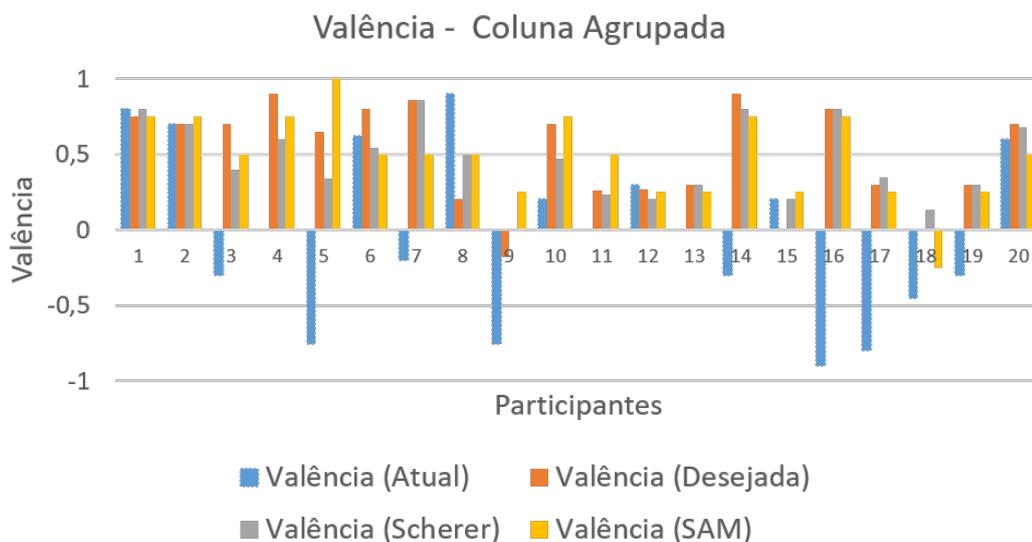
Para avaliar a emoção alcançada pelos participantes, além da análise das expressões faciais mencionadas anteriormente, foram aplicadas outras duas técnicas: o modelo auto-avaliação de emoção SAM representado pelas colunas em amarelo e o questionário subjetivo do espaço

**Figura 38 – Comparação dos níveis de excitação alcançada pelos participantes antes e após ouvir as músicas sugeridas pelo algoritmo.**



Fonte: Elaborada pelo autor.

**Figura 39 – Comparação dos níveis de valência alcançada pelos participantes antes e após ouvir as músicas sugeridas pelo algoritmo.**



Fonte: Elaborada pelo autor.

emocional semântico de Scherer representado pelas colunas em cinza.

As duas técnicas aplicadas no experimento mostraram que os 20 participantes distanciaram do nível da emoção atual e se aproximaram do estado emocional desejado. Além do mais, a avaliação pelo artefato de Scherer, mostra que dentre essa aproximação, 14 participantes emparelharam com alturas próximas aos níveis de excitação desejada e 12 com o nível da valência desejada. Já a técnica de SAM apresenta que 10 participantes emparelharam com o nível da excitação desejada e 11 para o nível da valência desejada.

**Tabela 4 – Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer.**

Participantes	Antes de ouvir	SAM	Scherer
1	0.07	0.10	0.07
2	0.85	0.11	0.13
3	1.08	0.28	0.30
4	0.91	0.38	0.30
5	1.64	0.43	0.35
6	0.90	0.34	0.37
7	1.42	0.37	0.20
8	0.71	0.30	0.34
9	1.01	0.51	0.33
10	0.52	0.21	0.35
11	0.33	0.51	0.04
12	0.50	0.55	0.12
13	0.58	0.25	0.10
14	1.20	0.34	0.13
15	0.54	0.25	0.28
16	1.71	0.07	0.00
17	1.36	0.07	0.11
18	0.64	0.25	0.16
19	0.68	0.05	0.12
20	0.51	0.21	0.30

**Fonte:** Elaborada pelo autor.

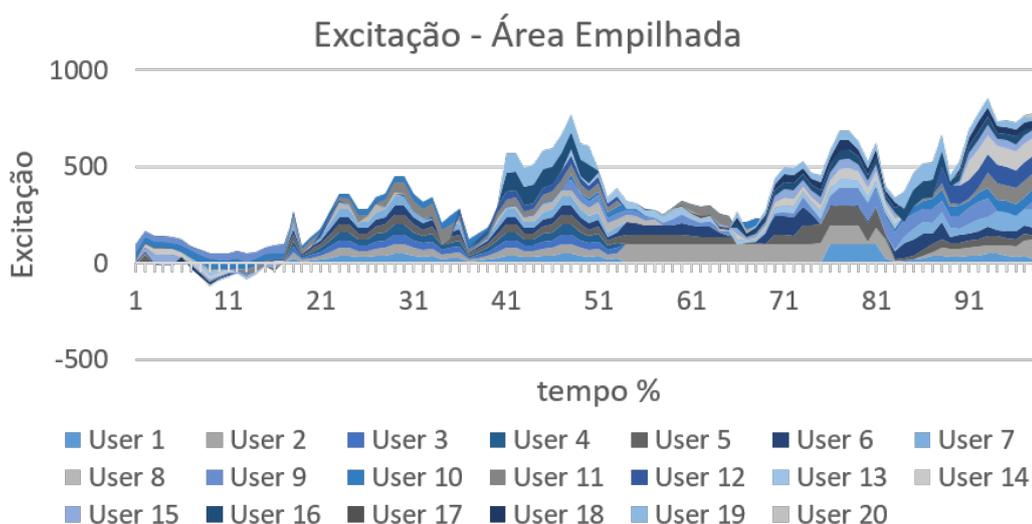
A Tabela 5.3.4 apresenta uma medida de distância entre a emoção desejada pelos participantes com as emoções obtidas por meio das avaliações utilizando o SAM, Scherer, bem como a emoção atual antes de ouvir as músicas. Por meio da tabela, é possível verificar que, se comparar os valores da distância antes de ouvir as músicas com as avaliadas com o SAM e o Scherer após a reprodução das músicas, percebe-se que quase todos os participantes se aproximaram da emoção desejada em ambas as avaliações aplicadas, com exceção dos participantes 1, 11 e 12 que tiveram uma distância maior apresentada pela avaliação de SAM.

De forma geral, as duas técnicas mostram que os usuários se aproximam da emoção desejada. Para cada participante foi calculado um percentual de ganho partindo da emoção atual até a emoção desejada. A partir dos percentuais de ganho atingidos por cada participante, uma média foi calculada entre todos os participantes resultando em um percentual de aproximação para as avaliações de Scherer e SAM. Com isso, tem-se uma média de ganho de aproximação para o estado emocional desejado de 53.7% para a técnica de SAM e 68.5% para o artefato de Scherer. Além do mais, a técnica de SAM apresenta que 85% dos participantes se aproximaram do estado emocional desejado e 95% dos participantes por meio do artefato de Scherer.

### 5.3.5 Resultados da Recomendação Manual do Participante

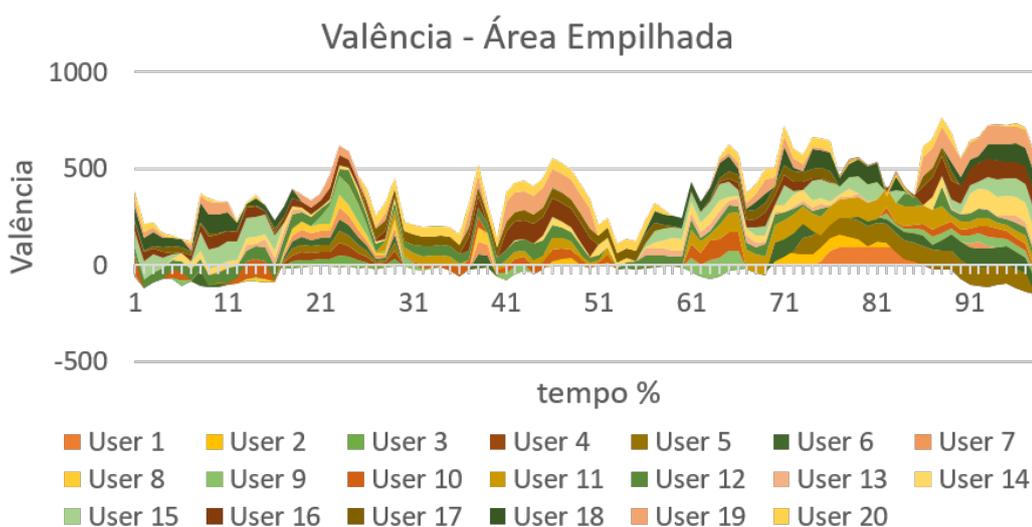
Durante a reprodução das músicas indicadas pelos participantes, as expressões faciais foram avaliadas, de maneira que a excitação e valência fossem recuperadas enquanto as músicas eram reproduzidas. As Figuras 40 e 41 mostram o gráfico da área empilhada no decorrer do tempo da reprodução das músicas para excitação e valência de cada participante.

**Figura 40 – Total acumulativo da excitação ao longo da reprodução das listas sugeridas para cada participante.**



**Fonte:** Elaborada pelo autor.

**Figura 41 – Total acumulativo da valência ao longo da reprodução das listas sugeridas para cada participante.**



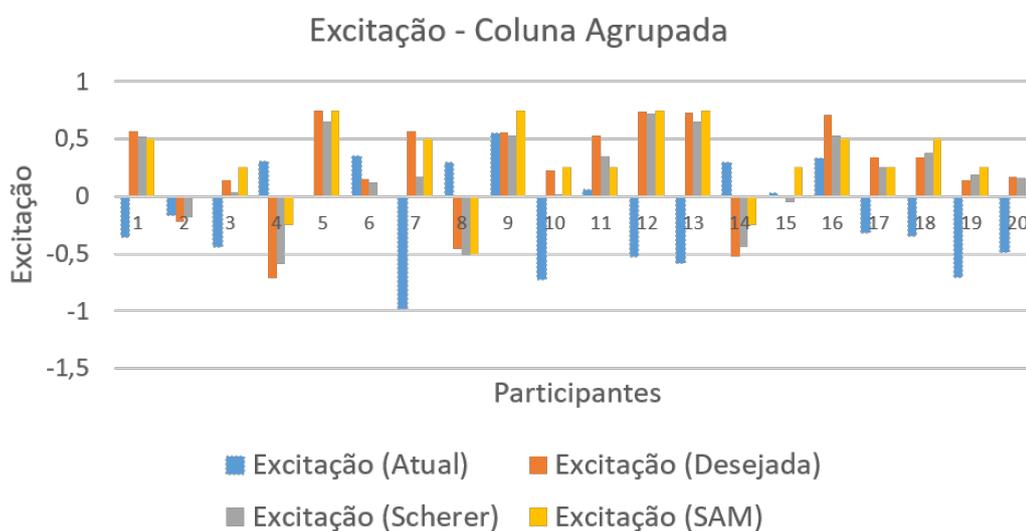
**Fonte:** Elaborada pelo autor.

O gráfico de Excitação da Figura 40 mostra que com a normalização do tamanho das listas de reprodução, a partir de 20% das listas os participantes tiveram uma variação da excitação

aumentada gradativamente, embora cada lista tivesse uma quantidade diferente de músicas. Diferente da Excitação, o gráfico da Valência na Figura 41 mostra que o aumento gradativo da variação acontece desde o início da reprodução da lista e que uma variação maior ocorre a partir de 60% das listas.

Após avaliar as emoções dos participantes durante a reprodução das músicas e verificar se produziam uma variação no estado emocional, verificou-se que após a escuta das músicas os participantes conseguiram se aproximar do estado emocional desejado. As Figuras 42 e 43 mostram os gráficos de excitação e valência para o estado emocional atual, desejado e alcançado de cada participante.

**Figura 42 – Comparação dos níveis de excitação alcançada pelos participantes antes e após ouvir as músicas sugeridas pelo algoritmo.**



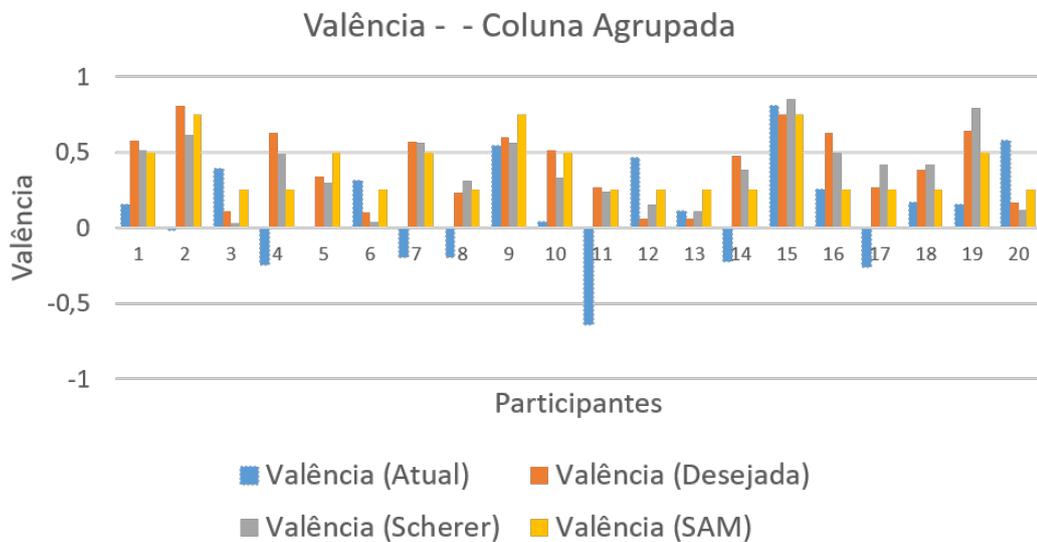
**Fonte:** Elaborada pelo autor.

Para avaliar a emoção alcançada pelos participantes, além da análise das expressões faciais, foram aplicadas duas técnicas: o modelo auto-avaliação de emoção SAM representado pelas colunas em amarelo e o questionário subjetivo do espaço emocional semântico de Scherer representado pelas colunas em cinza.

As duas técnicas aplicadas no experimento mostraram que os 20 participantes se aproximaram do estado emocional desejado. Além do mais, a avaliação pelo artefato de Scherer, mostra que dentre essa aproximação, 17 participantes emparelharam com alturas próximas ao nível da excitação desejada e 17 com o nível da valência desejada. Já a técnica de SAM apresenta que 12 participantes emparelharam com o nível da excitação desejada e 13 para o nível da valência desejada.

A Tabela 5.3.5 apresenta uma medida de distância entre a emoção desejada pelos participantes com as emoções obtidas por meio das avaliações utilizando o SAM, Scherer, bem como a emoção atual antes de ouvir as músicas. Por meio da tabela, é possível verificar que, se ao

**Figura 43 – Comparação dos níveis de valência alcançada pelos participantes antes e após ouvir as músicas sugeridas pelo algoritmo.**



**Fonte:** Elaborada pelo autor.

comparar os valores da distância antes de ouvir as músicas com as avaliadas com o SAM e o Scherer após a reprodução das músicas, percebe-se que quase todos os participantes aproximaram da emoção desejada em ambas as avaliações aplicadas, com exceção o participante 15 que teve uma distância maior do que antes de ouvir.

De modo geral, é possível perceber que as duas técnicas mostram que os usuários se aproximam da emoção desejada, tendo como média um ganho de aproximação para o estado emocional desejado de 38% apresentado pela técnica de SAM e 72.5% para o artefato de Scherer. Além do mais, 90% dos participantes aproximaram do estado emocional desejado apresentado pelo SAM e 95% por meio do Scherer.

## 5.4 Recomendação pelo Reprodutor m-Motion

### 5.4.1 Emoção Atual e Desejada do Usuário

Durante o experimento, cada participante teve sua emoção atual obtida por dois artefatos: pelo texto da voz, no qual o participante discute como está se sentindo no instante da utilização do reprodutor e por meio do componente visual de Scherer.

A emoção desejada era obtida apenas pelo componente de Scherer. A atividade consistia em marcar um ponto no componente de Scherer, de tal forma que representasse a emoção que se desejava alcançar.

**Tabela 5 – Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer.**

Participantes	Antes de ouvir	SAM	Scherer
1	1.02	0.11	0.09
2	0.82	0.23	0.20
3	0.64	0.18	0.13
4	1.33	0.60	0.18
5	0.82	0.16	0.11
6	0.29	0.21	0.07
7	1.73	0.10	0.40
8	0.86	0.04	0.09
9	0.06	0.24	0.05
10	1.06	0.02	0.28
11	1.02	0.28	0.18
12	1.32	0.19	0.09
13	1.31	0.19	0.09
14	1.07	0.35	0.13
15	0.07	0.25	0.11
16	0.54	0.43	0.22
17	0.84	0.09	0.17
18	0.71	0.21	0.06
19	0.97	0.18	0.16
20	0.77	0.19	0.05

**Fonte:** Elaborada pelo autor.

### 5.4.2 Reconhecimento Emocional da Música

Após capturar a emoção atual e desejada dos participantes por meio dos componentes, os metadados das músicas dos participantes eram recuperados do servidor do LIFES. Cada participante tinha a sua lista de músicas recuperada separadamente e com as devidas emoções classificadas pelo servidor do LIFES. Portanto, cada participante era guiado para a tela da lista de reprodução. Após receber a lista de sugeridas pelo algoritmo de recomendação do m-Motion, cada participante era instruído a reproduzir todas as músicas na ordem sugerida antes de realizar a etapa de avaliação da recomendação. A reprodução das músicas poderia ser realizada por qualquer plataforma de streaming de música de preferência do usuário, desde que seguisse a ordem sugerida pela aplicação m-Motion. Além do mais, a aplicação fornecia para cada música um link para reprodução na plataforma YouTube<sup>2</sup>.

### 5.4.3 Avaliação da Emoção Alcançada

Ao fim da reprodução das músicas, cada participante foi instruído a realizar a avaliação da lista recomendada. A avaliação consistia em verificar a emoção alcançada pelos participantes

<sup>2</sup> YouTube - Plataforma de compartilhamento de vídeos.

após a reprodução das músicas. Para verificar o estado emocional dos participantes após a reprodução das músicas e obter o valor de excitação e valência da emoção correspondente, foi utilizado o componente de Scherer e o componente de SAM. Em seguida, as informações eram enviadas ao servidor do LIFES para a análise das recomendações do m-Motion.

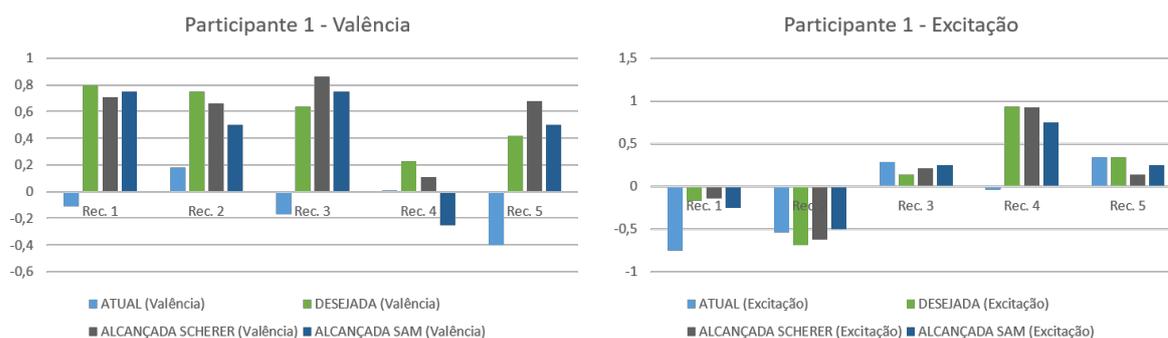
#### 5.4.4 Resultados da Recomendação pela Aplicação m-Motion

O recrutamento dos participantes foi realizado por conveniência, considerando os requisitos técnicos de implementação e implantação da aplicação em seus smartphones. Portanto, 8 participantes atenderam esses critérios, sendo que era necessário um smartphone com sistema operacional iOS sujeito a instalação da aplicação m-Motion via computador e execução da aplicação durante 5 dias em uma semana. Em seguida é apresentado os resultados obtidos após a realização dos experimentos com o uso real da aplicação m-Motion com 8 participantes. Para cada participante é apresentado um gráfico de excitação e valência que representa os níveis alcançados por cada participante antes e após a reprodução das músicas.

Os gráficos possuem 4 colunas para cada recomendação realizada pelos participantes. Cada coluna referencia os estados emocionais dos participantes, sendo que, as colunas em azul-claro representam a emoção atual antes da reprodução das músicas, as colunas em verde a emoção desejada, as colunas em cinza representa a emoção alcançada por meio do componente de Scherer e em azul-escuro a emoção alcançada por meio do componente de SAM.

A Figura 44 apresenta os gráficos dos níveis de valência e excitação obtidos do participante 1 durante os experimentos. O gráfico da valência mostra que as recomendações 1, 2, 3 e 5, o participante aproximou da emoção desejada que é representada pela coluna em verde. Além do mais, é possível perceber que a recomendação 4 o participante se aproximou da emoção desejada apenas na avaliação realizada pelo componente de Scherer e que a avaliação de SAM se distanciou da emoção desejada. Já no gráfico da excitação, em todas as recomendações são possíveis perceber que o nível da emoção alcançada pelos dois métodos de avaliação se aproximaram da emoção desejada e se distanciaram da emoção atual.

**Figura 44 – Participante 1 - Níveis de Valência e Excitação.**



**Fonte:** Elaborada pelo autor.

A Tabela 5.4.4 apresenta uma medida de distância entre a emoção desejada pelos participantes com as emoções obtidas por meio das avaliações utilizando o componente de Scherer e SAM, bem como a emoção atual antes de ouvir as músicas. Por meio da tabela, é possível verificar que, se ao comparar os valores da distância antes de ouvir as músicas com as avaliadas com o SAM e o Scherer após a reprodução das músicas, percebe-se que as recomendações realizadas pelo algoritmo do m-Motion ajudaram os participantes se aproximarem da emoção desejada em ambas avaliações realizadas pelos componentes.

De maneira geral, as duas técnicas mostram que o participante aproxima da emoção desejada, tendo como média de um ganho de aproximação para o estado emocional desejado de 70,5% apresentado pela técnica de SAM e 78,3% para o artefato de Scherer.

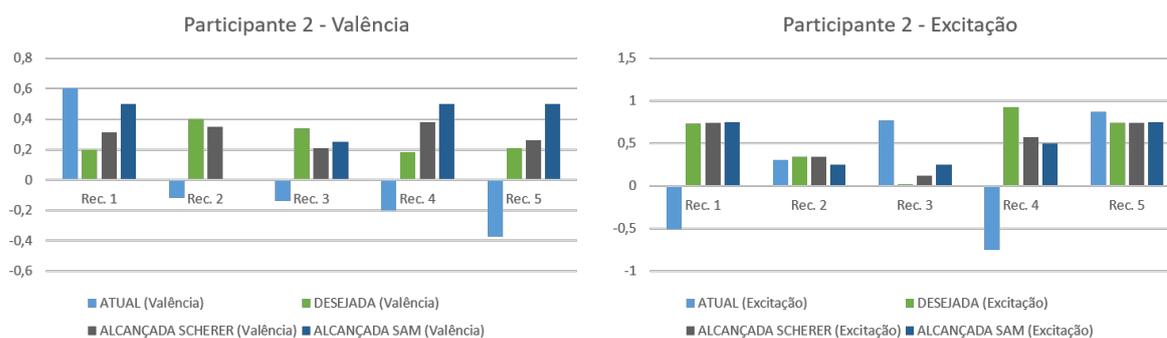
**Tabela 6 – Participante 1 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer.**

Recomendações	Antes de ouvir	Scherer	SAM
Rec. 1	1.08	0.09	0.09
Rec. 2	0.59	0.11	0.31
Rec. 3	0.82	0.23	0.16
Rec. 4	0.99	0.12	0.51
Rec. 5	0.82	0.33	0.12

Fonte: Elaborada pelo autor.

O participante 2 contém os níveis de valência e excitação recuperados durante o experimento apresentado na Figura 45. O gráfico da valência mostra que para recomendações 1, 3, 4 e 5, a emoção alcançada pelo participante tendeu-se aproximar da emoção desejada em ambas avaliações. No entanto, para a recomendação 2, apenas a avaliação do artefato de Scherer mostra que o participante se aproxima da emoção desejada. Já no gráfico da excitação, em todas as recomendações são possíveis perceber que o nível da emoção alcançada pelos dois métodos de avaliação se aproximaram da emoção desejada e se distanciaram da emoção atual.

**Figura 45 – Participante 2 - Níveis de Valência e Excitação.**



Fonte: Elaborada pelo autor.

Por meio da Tabela 5.4.4 é possível perceber que, ao comparar os valores da distância antes de ouvir as músicas com as avaliadas com o SAM e o Scherer após a reprodução das músicas, percebe-se que as recomendações realizadas pelo algoritmo do m-Motion ajudou o participante aproximar da emoção desejada em ambas avaliações realizadas pelos componentes. Além do mais, as avaliações do componente de Scherer apresentou distâncias inferiores a SAM para todas as recomendações, isso porque a avaliação de Scherer retornou uma distância menor da emoção desejada. As duas técnicas mostram que o participante se aproxima da emoção desejada, tendo como média de um ganho de aproximação para o estado emocional desejado de 58,1% apresentado pela técnica de SAM e 86,3% para o artefato de Scherer.

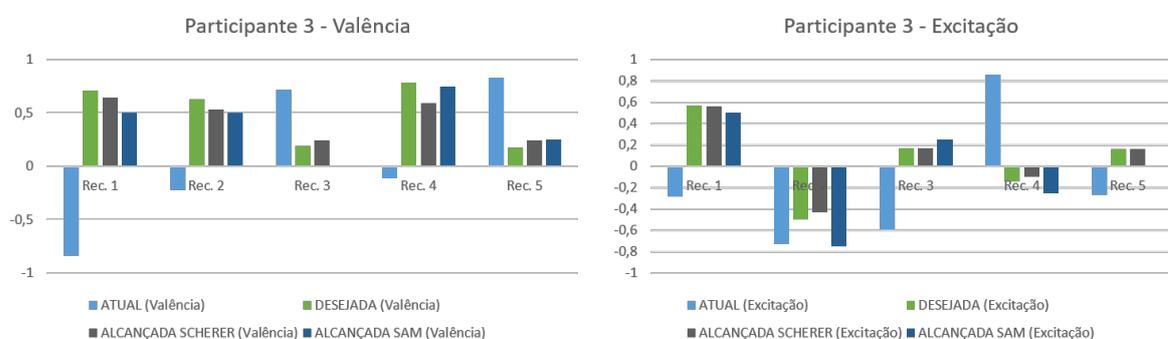
**Tabela 7 – Participante 2 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer.**

Recomendações	Antes de ouvir	Scherer	SAM
Rec. 1	1.30	0.11	0.30
Rec. 2	0.52	0.05	0.41
Rec. 3	0.89	0.16	0.25
Rec. 4	1.71	0.40	0.53
Rec. 5	0.59	0.05	0.29

**Fonte:** Elaborada pelo autor.

Os níveis de valência e excitação do participante 3 recuperados durante o experimento são apresentados na Figura 46. O gráfico da valência mostra que para todas as recomendações realizadas pelo algoritmo da aplicação m-Motion, a emoção alcançada pelo participante tendeu-se aproximar da emoção desejada. Já no gráfico da excitação, as recomendações 1, 3, 4 e 5 teve o nível de excitação alcançado próximo ao desejado para ambas avaliações. Além do mais, é possível perceber que na recomendação 2, apenas a avaliação utilizando o componente de Scherer apresentou que o participante se aproximou do nível da excitação desejada.

**Figura 46 – Participante 3 - Níveis de Valência e Excitação.**



**Fonte:** Elaborada pelo autor.

A Tabela 5.4.4 mostra que é as recomendações realizadas pelo algoritmo do m-Motion ajudou o participante 3 a se aproximar da emoção desejada em ambas avaliações realizadas pelos

componentes. Além do mais, as avaliações de Scherer para o participante 3 apresentou distâncias menores que a avaliação de SAM para as recomendações 1, 2, 3 e 5 e, em contrapartida, a avaliação de SAM apresentou distância menor na recomendação 4. As duas técnicas mostram que o participante se aproxima da emoção desejada, tendo como média de ganho de aproximação para o estado emocional desejado de 80,4% apresentado pela técnica de SAM e 90,9% para o artefato de Scherer.

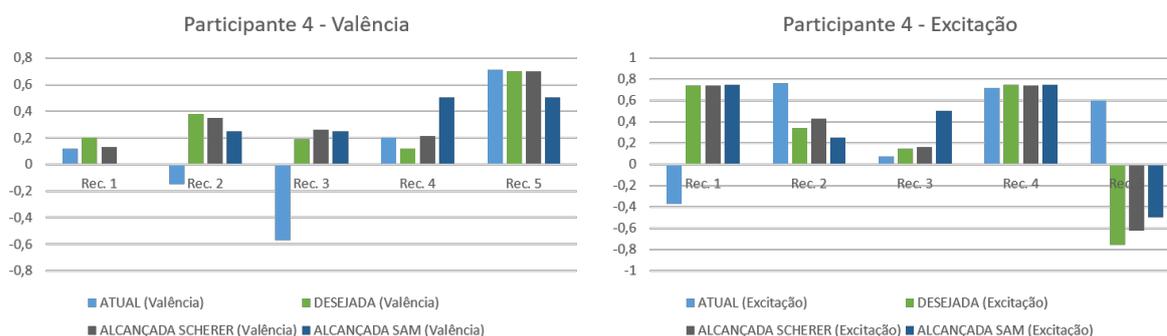
**Tabela 8 – Participante 3 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer.**

Recomendações	Antes de ouvir	Scherer	SAM
Rec. 1	1.77	0.07	0.22
Rec. 2	0.88	0.12	0.28
Rec. 3	0.93	0.05	0.21
Rec. 4	1.34	0.19	0.11
Rec. 5	0.78	0.06	0.17

Fonte: Elaborada pelo autor.

A Figura 47 apresenta os gráficos dos níveis de valência e excitação obtidos do participante 4 durante os experimentos. O gráfico da valência mostra que as recomendações 1, 2, 3 e 5, o participante se aproximou da emoção desejada em ambas avaliações. Já no gráfico da excitação, as recomendações 1, 2, 4 e 5 o nível da emoção alcançada pelos dois métodos de avaliação se aproximaram da emoção desejada, porém, na recomendação 3, a emoção recuperada pela avaliação de SAM se distanciou da emoção desejada.

**Figura 47 – Participante 4 - Níveis de Valência e Excitação.**



Fonte: Elaborada pelo autor.

Por meio da Tabela 5.4.4, é possível verificar que, se comparar os valores da distância antes de ouvir as músicas com as avaliadas com o SAM e o Scherer após a reprodução das músicas, percebe-se que as recomendações realizadas pelo algoritmo do m-Motion ajudou o participante 4 se aproximar da emoção desejada nas recomendações 1, 2, 3 e 5 em ambas avaliações realizadas pelos componentes. Além do mais, as avaliações do componente de Scherer

apresentou distâncias menores que a avaliação de SAM para todas as recomendações. É possível perceber também, que na recomendação 4, ambas avaliações mostraram que o participante não se aproximou da emoção desejada, sendo que na avaliação de Scherer o participante manteve a emoção atual e na avaliação de SAM a distância foi maior que antes de ouvir as músicas sugeridas.

Com exceção a recomendação 4, o participante se aproxima da emoção desejada para as recomendações 1, 2, 3 e 5, tendo como média um ganho de aproximação para o estado emocional desejado de 70,8% para o artefato de Scherer. No entanto, para a técnica de SAM a média foi de -11,3%. Isso ocorre pelo fato da recomendação 4 distanciar o participante da emoção desejada. Ressalta-se também, que apenas na recomendação 4 as avaliações mostraram que o participante não alcançou o estado emocional desejado, no entanto, as demais recomendações as avaliações mostram que o participante conseguiu-se aproximar da emoção desejada. Entende-se que tal divergência pode ter ocorrido devido ao fato do participante buscar alcançar uma emoção próxima a atual, visto que nas demais recomendações o participante não buscava alcançar um estado emocional próximo a atual.

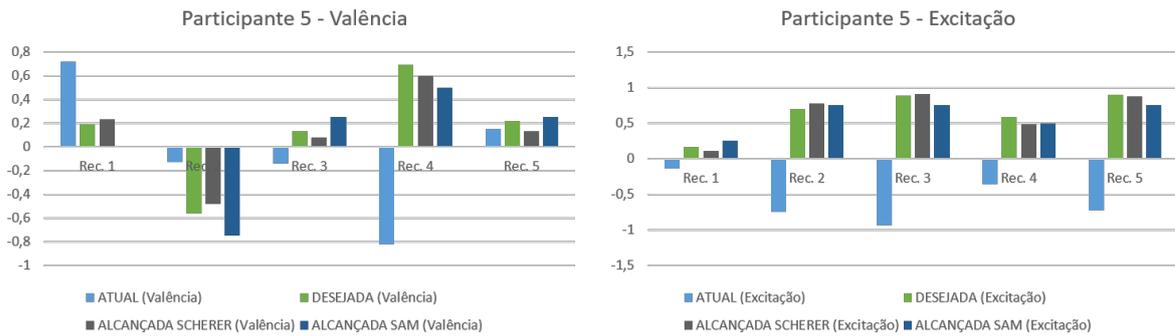
**Tabela 9 – Participante 4 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer.**

Recomendações	Antes de ouvir	Scherer	SAM
Rec. 1	1.11	0.07	0.20
Rec. 2	0.68	0.09	0.16
Rec. 3	0.76	0.07	0.36
Rec. 4	0.09	0.09	0.38
Rec. 5	1.36	0.14	0.33

**Fonte:** Elaborada pelo autor.

Os níveis de excitação e valência alcançados pelo participante 5 apresentados na Figura 48, mostram que após ouvir as músicas recomendadas pela aplicação m-Motion, o participante 5 se aproximou da emoção desejada em ambas avaliações. Contudo, a Tabela 5.4.4 é apresentada para exibir as distâncias entre a emoção desejada e alcançada pelo participante.

**Figura 48 – Participante 5 - Níveis de Valência e Excitação.**



**Fonte:** Elaborada pelo autor.

A Tabela 5.4.4 mostra que nas duas avaliações realizadas, o participante diminuiu a distância entre a emoção atual e a emoção desejada. Além do mais, a avaliação do componente de Scherer resultou em uma distância menor para todas as recomendações se comparadas com a avaliação de SAM. As duas técnicas mostram que o participante se aproximou da emoção desejada, tendo como média de um ganho de aproximação para o estado emocional desejado de 84,4% apresentado pela avaliação de SAM e 92,8% para a avaliação utilizando o componente de Scherer.

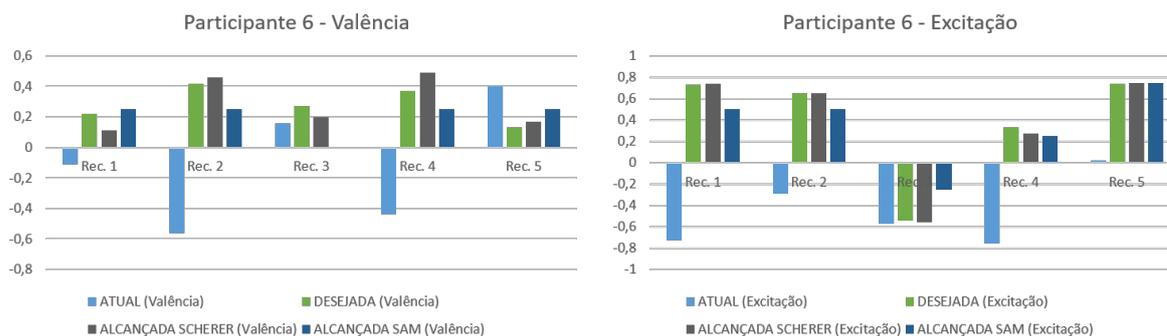
**Tabela 10 – Participante 5 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer.**

Recomendações	Antes de ouvir	Scherer	SAM
Rec. 1	0.61	0.07	0.21
Rec. 2	1.51	0.11	0.20
Rec. 3	1.85	0.05	0.18
Rec. 4	1.78	0.13	0.21
Rec. 5	1.62	0.09	0.15

**Fonte:** Elaborada pelo autor.

A Figura 49 mostra os gráficos dos níveis de valência e excitação recuperados do participante 6 durante o experimento. O gráfico da valência mostra que as recomendações 1, 2, 4 e 5, o participante se aproximou da emoção desejada. Além do mais, é possível perceber que na recomendação 3 a avaliação realizada pelo componente de Scherer o participante se aproximou no nível da emoção desejada e que na avaliação de SAM o participante se distanciou do nível da emoção desejada. Já no gráfico da excitação, as recomendações 1, 2, 4 e 5 o nível da emoção alcançada pelos dois métodos de avaliação se aproximaram da emoção desejada, porém, na recomendação 3, outra vez a avaliação de SAM mostrou que o participante não se aproximou do nível da emoção desejada.

**Figura 49 – Participante 6 - Níveis de Valência e Excitação.**



**Fonte:** Elaborada pelo autor.

Quando o nível da excitação ou valência do participante não se aproxima do nível desejado, por consequência isso reflete na distância entre a emoção alcançada com a desejada, e isso pode ser visto na Tabela 5.4.4. A avaliação de SAM mostra que o nível de valência e excitação alcançados na recomendação 3 não foram próximos aos níveis desejados, portanto, isso reflete na distância da recomendação 3 apresentada na tabela. Como pode ser visto na tabela, a avaliação de SAM para a recomendação 3 apresentou uma distância maior que a distância da avaliação de Scherer.

De modo geral, o participante 6 se aproxima da emoção desejada, tendo como média de ganho uma aproximação para o estado emocional desejado de 82,11% para o artefato de Scherer. No entanto, para a técnica de SAM a média foi de 18,8%. Isso ocorre pelo fato da recomendação 3 distanciar o participante da emoção desejada. Entende-se que tal divergência pode ter ocorrido devido ao fato do participante buscar alcançar uma emoção próxima a atual, visto que nas demais recomendações o participante não buscava alcançar um estado emocional próximo a atual.

**Tabela 11 – Participante 6 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer.**

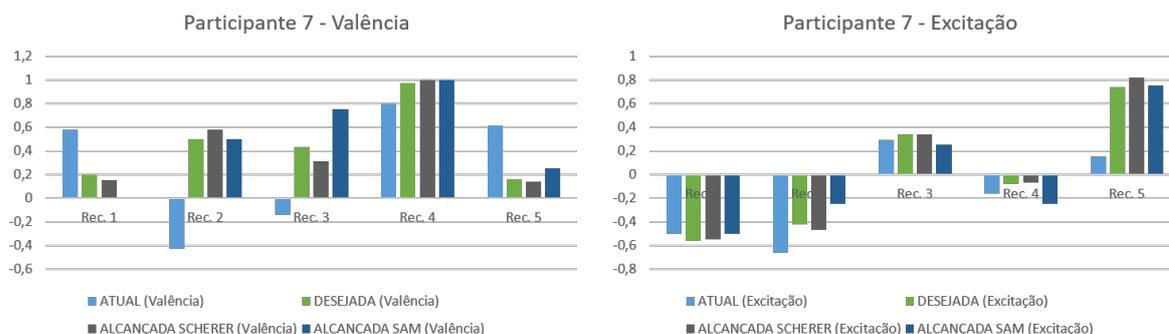
Recomendações	Antes de ouvir	Scherer	SAM
Rec. 1	1.50	0.11	0.23
Rec. 2	1.36	0.04	0.23
Rec. 3	0.11	0.07	0.40
Rec. 4	1.36	0.13	0.14
Rec. 5	0.77	0.04	0.12

**Fonte:** Elaborada pelo autor.

O participante 7 contém os níveis de valência e excitação recuperados durante o experimento apresentado na Figura 50. O gráfico da valência mostra que para recomendações 1, 2, 4 e 5, a emoção alcançada pelo participante tendeu-se aproximar da emoção desejada. No entanto, é possível perceber que na recomendação 3 apenas a avaliação realizada pelo componente de Scherer mostrou que o participante se aproximou da emoção desejada e que a avaliação de

SAM o participante se distanciou da emoção desejada. O gráfico da excitação mostra que as recomendações 1, 2, 3 e 5 a emoção alcançada em ambas avaliações se aproximaram da emoção desejada, porém, a recomendação 4 isso não ocorreu para ambos, mas apenas para a avaliação de Scherer.

**Figura 50 – Participante 7 - Níveis de Valência e Excitação.**



**Fonte:** Elaborada pelo autor.

A Tabela 5.4.4 mostra que ambas avaliações levaram o participante 7 se aproximar da emoção desejada, isso se explica, pois, a distância entre a emoção alcançada com a desejada diminuiu se comparadas com a distância da emoção antes de ouvir as músicas. Além do mais a avaliação de Scherer apresentou distâncias menores que a avaliação de SAM para todas as recomendações. As duas técnicas mostram que o participante se aproximou da emoção desejada, tendo como média um ganho de aproximação para o estado emocional desejado de 53,1% apresentado pela técnica de SAM e 85,6% para o artefato de Scherer.

**Tabela 12 – Participante 7 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer.**

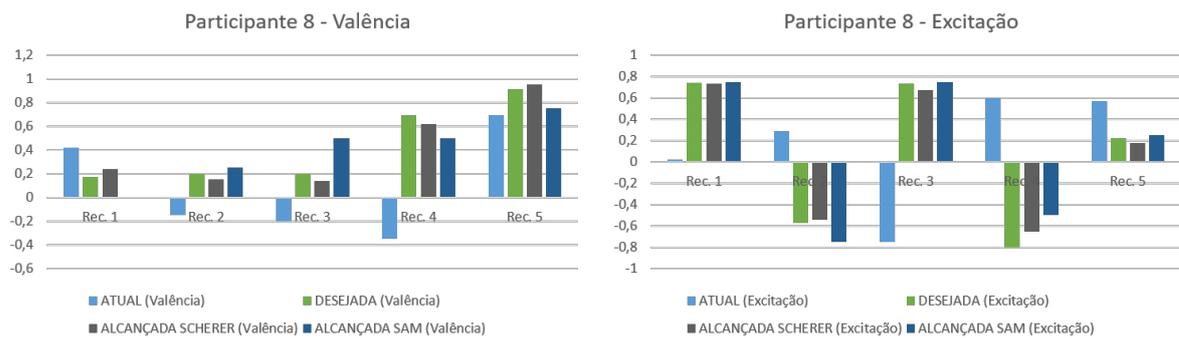
Recomendações	Antes de ouvir	Scherer	SAM
Rec. 1	0.38	0.05	0.21
Rec. 2	0.96	0.09	0.17
Rec. 3	0.57	0.12	0.33
Rec. 4	0.19	0.03	0.17
Rec. 5	0.74	0.08	0.09

**Fonte:** Elaborada pelo autor.

Os níveis de excitação e valência alcançados pelo participante 8 apresentados na Figura 48, mostram que após ouvir as músicas recomendadas pela aplicação m-Motion, o participante 8 se aproximou da emoção desejada em ambas avaliações para as recomendações 1, 2, 4 e 5. Na recomendação 3, apenas a avaliação de Scherer apresentou que o participante 8 se aproximou da emoção desejada. Em contrapartida, no gráfico da excitação, ambas avaliações mostram que a emoção alcançada foi aproximada da emoção desejada pelo participante 8.

Por meio da Tabela 5.4.4 é possível verificar que a distância entre a emoção alcançada com a desejada diminuiu se comparadas com a distância da emoção antes de ouvir as músicas. Além do mais a avaliação de Scherer apresentou distâncias menores para todas as recomendações se comparadas com a avaliação de SAM. O participante se aproximou da emoção desejada, tendo como média um ganho de aproximação para o estado emocional desejado de 75,6% para a técnica de SAM e 91,1% para o artefato de Scherer.

**Figura 51 – Participante 8 - Níveis de Valência e Excitação.**



**Fonte:** Elaborada pelo autor.

**Tabela 13 – Participante 8 - Comparação da distância entre a emoção desejada e as emoções obtidas através das avaliações de SAM e Scherer.**

Recomendações	Antes de ouvir	Scherer	SAM
Rec. 1	0.76	0.07	0.17
Rec. 2	0.93	0.06	0.19
Rec. 3	1.53	0.08	0.30
Rec. 4	1.74	0.17	0.36
Rec. 5	0.41	0.06	0.16

**Fonte:** Elaborada pelo autor.

# Capítulo 6

## CONCLUSÕES

---

---

### 6.1 Análise Crítica

Os resultados obtidos sugerem que, ao ouvir música selecionada pelo algoritmo proposto, os usuários podem se aproximar do estado emocional desejado. No entanto, vários aspectos podem influenciar diretamente a recomendação das músicas, por exemplo, a precisão na recuperação da emoção atual do usuário e a classificação da emoção das músicas.

O algoritmo recebe 3 entradas de parâmetros para em seguida inicializar o processo de recomendação das músicas, sendo como primeiro parâmetro a emoção atual do usuário, o segundo a emoção desejada e o terceiro uma lista de músicas classificadas em excitação e valência.

Nos três experimentos realizados, a emoção atual dos participantes foram recuperadas por meio do artefato do espaço semântico de Scherer, que por decorrência resultavam em valores de valência e excitação. No entanto, a emoção atual pode ser recuperada com mais precisão se vinda por sensores ou uma fusão de várias entradas. Uma única entrada pode não ser tão precisa para detectar a emoção atual do usuário. Além do mais o usuário pode não saber exatamente o estado emocional em que se encontra.

A segunda entrada é a emoção desejada, que também foi obtida por uma marcação no espaço semântico. Outras formas de captura também podem ser aplicadas, como, por exemplo, a utilização de imagens que representam estados emocionais em que os usuários gostariam de alcançar.

Nos experimentos em que os participantes sugeriram suas próprias músicas ou receberam recomendações pelo algoritmo, a análise das expressões faciais foi usada para monitorar as variações nas emoções dos participantes e para verificar se eles aproximavam do estado emocional desejado. No entanto, notou-se que a maioria dos participantes expressou seu entusiasmo pela música através de movimentos com as mãos, pés e cabeça, o que não pôde ser percebido apenas com a análise das expressões faciais.

Uma das entradas do algoritmo de recomendação é a lista de músicas do usuário com suas

emoções classificadas. Portanto, espera-se que o classificador de música tenha uma alta precisão quanto à sua classificação. Contudo, se o algoritmo recebe a lista de músicas não classificadas corretamente, por mais que o algoritmo realize todo o processo de encontrar as músicas partindo do ponto da emoção atual até a emoção desejada, as músicas sugeridas não poderão surgir os efeitos esperados nos usuários devido à classificação incorreta. Ou seja, o algoritmo se mostra eficiente em realizar todo o processo de recuperação e recomendação de músicas para levar o usuário a um estado emocional desejado, desde que o modelo de classificação da emoção das músicas seja preciso e eficiente.

Se as músicas não são classificadas corretamente, por mais que o algoritmo proposto realize todo o processo de recuperação das músicas do ponto da emoção atual para a emoção desejada, as músicas sugeridas não poderão surgir os efeitos esperados nos usuários.

As avaliações realizadas nos experimentos de recomendação manual mostram que 90% dos participantes se aproximaram da emoção desejada por meio da técnica de SAM e 95% por meio do artefato de Scherer.

Em contrapartida, na recomendação pelo algoritmo, a avaliação por meio da técnica de SAM mostrou que 85% dos participantes se aproximaram da emoção desejada e 95% pelo artefato de Scherer. É possível perceber também, que nos gráficos das expressões faciais, se comparados os dois métodos de recomendação, a recomendação manual superou a recomendação pelo algoritmo. No entanto, esse resultado era esperado e foi confirmado com os experimentos, visto que o usuário conhece a sua lista de música e é apto a julgar quais músicas o levam para um estado emocional desejado.

Contudo, o objetivo do experimento de recomendação pelo algoritmo, não era superar a forma manual em que os usuários realizam no seu dia a dia, mas sim mostrar que a recomendação do algoritmo se aproxima do método manual realizado pelo usuário. Na forma manual, os usuários possuem uma variação do estado emocional a partir de 20% da reprodução da lista de músicas para a excitação enquanto para a valência há uma variação contínua desde o início da reprodução. Já para a recomendação feita pelo algoritmo a variação para a excitação e valência ocorre após 40% da reprodução das músicas.

Ao verificar que a recomendação feita pelo algoritmo se aproximou da recomendação manual, o terceiro experimento foi realizado. O terceiro experimento tinha como objetivo utilizar uma aplicação que pudesse ser realizada em tempo real e fora de ambiente controlado com o uso do algoritmo de recomendação. Sendo assim a aplicação m-Motion foi construída e os participantes puderam participar da execução do terceiro experimento.

No terceiro experimento, observa-se que ao longo da avaliação dos 8 participantes, a avaliação realizada pelo componente de Scherer apresentou as medidas de distância entre a emoção desejada e alcançada, valores menores do que as apresentadas na avaliação do componente de SAM. No total foram 40 recomendações realizadas pelo algoritmo do m-Motion,

sendo que das 36 recomendações a avaliação de Scherer apresentou distâncias menores que a avaliação de SAM. Além do mais, é possível perceber que quando os participantes buscavam alcançar um estado emocional próximo do atual a avaliação de SAM apontava que o usuário não se aproximava do estado emocional desejado e isso trouxe valores de distância maiores que as distâncias antes de ouvir as músicas.

## 6.2 Publicações

Este trabalho de dissertação apresenta 3 experimentos distintos, sendo: a avaliação da recomendação do algoritmo proposto, uma avaliação da recomendação manual do usuário e uma avaliação da aplicação m-motion em ambiente real.

Após a realização dos experimentos de recomendação pelo algoritmo terem sido realizados, um artigo foi produzido para apresentar a proposta científica do algoritmo. O trabalho cujo título é "An algorithm for music recommendation based on the user's musical preferences and desired emotions" foi submetido e aprovado pela conferência "International Conference on Mobile and Ubiquitous Multimedia", conforme é apresentado na Tabela 14.

**Tabela 14 – Meios de publicação**

<b>SIGLA</b>	<b>NOME DO EVENTO</b>	<b>QUALIS</b>
MUM	ACM 17th International Conference on Mobile and Ubiquitous Multimedia	B1

A conferência ocorreu no ano de 2018 na cidade de Cairo, capital do Egito, durante os dias 25 e 28 de Novembro. O MUM é um evento reconhecido pela CAPES com qualis B1.

## 6.3 Trabalhos Futuros

No futuro, planeja-se realizar um experimento em que a emoção atual do usuário possa ser recuperada a partir de várias entradas de sensores, bem como avaliar as reações dos usuários durante a reprodução das músicas. Além disso, deseja-se melhorar a avaliação e deixar os usuários criarem suas próprias listas de músicas que acreditam deixá-los em um estado emocional desejado e comparamos com a lista retornada pelo nosso algoritmo.

Há também um interesse de considerar o contexto do usuário. Aprimorar a aplicação m-Motion com o uso de redes neurais para que se possa aprender com os gostos do usuário bem como as suas atividades realizadas. A aplicação poderá verificar o estado atual do usuário por meio de sensores fisiológicos, assim como, o tempo, movimentação e outros dados externos. Com isso, o sistema poderá aprender a prever as músicas que o usuário gostaria de ouvir em um determinando momento para alcançar um estado emocional desejado.

## REFERÊNCIAS

---

- AGARWAL, A.; MEYER, A. Beyond Usability: Evaluating Emotional Response As an Integral Part of the User Experience. In: *CHI '09 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2009. (CHI EA '09), p. 2919–2930. ISBN 978-1-60558-247-4. Citado na página 21.
- BARRETT, L. F.; WAGER, T. D. The structure of emotion: Evidence from neuroimaging studies. *Current Directions in Psychological Science*, v. 15, n. 2, p. 79–83, 2006. Disponível em: <<https://doi.org/10.1111/j.0963-7214.2006.00411.x>>. Citado 2 vezes nas páginas 22 e 23.
- BARTOSZEWSKI, M.; KWASNICKA, H.; MARKOWSKA-KACZMAR, U.; MYSZKOWSKI, P. B. Extraction of emotional content from music data. In: IEEE. *Computer Information Systems and Industrial Management Applications, 2008. CISIM'08. 7th*. [S.l.], 2008. p. 293–299. Citado 2 vezes nas páginas 25 e 27.
- BENNETT, R. *History of Music*. Cambridge University Press, 1987. (Cambridge Assignments in Music). ISBN 9780521336819. Disponível em: <<https://books.google.com.br/books?id=GoZrQgAACAAJ>>. Citado na página 24.
- BRADLEY, M. M.; LANG, P. J. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, Elsevier, v. 25, n. 1, p. 49–59, 1994. Citado 4 vezes nas páginas 19, 30, 31 e 71.
- BRITANNICA, E. *Music*. Encyclopædia Britannica, inc., 2016. Disponível em: <<https://www.britannica.com/art/music>>. Citado na página 16.
- CERVANTES, C. A.; SONG, K. T. Embedded design of an emotion-aware music player. *Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013*, p. 2528–2533, 2013. ISSN 1062-922X. Citado 6 vezes nas páginas 17, 32, 38, 39, 40 e 46.
- CHAPANERI, S.; LOPES, R.; JAYASWAL, D. Evaluation of music features for puk kernel based genre classification. *Procedia Computer Science*, Elsevier, v. 45, p. 186–196, 2015. Citado na página 29.
- CHIN, Y.-H.; LIN, P.-C.; TAI, T.-C.; WANG, J.-C. Genre based emotion annotation for music in noisy environment. In: IEEE. *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*. [S.l.], 2015. p. 863–866. Citado na página 29.
- DENG, J. J.; LEUNG, C. Emotion-based music recommendation using audio features and user playlist. In: *2012 6th International Conference on New Trends in Information Science, Service Science and Data Mining (ISSDM2012)*. [S.l.: s.n.], 2012. p. 796–801. Citado 4 vezes nas páginas 18, 32, 42 e 47.

DESMET, P. et al. Measuring emotion: development and application of an instrument to measure emotional responses to products. *Funology: from usability to enjoyment*, p. 111–123, 2003. Citado na página 22.

DOBASHI, A.; IKEMIYA, Y.; ITOYAMA, K.; YOSHII, K. A music performance assistance system based on vocal, harmonic, and percussive source separation and content visualization for music audio signals. *Proceedings of SMC*, p. 99–104, 2015. Citado na página 25.

DORNBUSH, S.; FISHER, K.; MCKAY, K.; PRIKHODKO, A.; SEGALL, Z. XPOD - A Human Activity and Emotion Aware Mobile Music Player. *2005 2nd Asia Pacific Conference on Mobile Technology, Applications and Systems*, 2005. Citado 7 vezes nas páginas 17, 32, 33, 34, 45, 46 e 48.

FAN, S.; TAN, C.; FAN, X.; SU, H.; ZHANG, J. HeartPlayer: A smart music player involving emotion recognition, expression and recommendation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, v. 6524 LNCS, n. PART 2, p. 483–485, 2011. ISSN 03029743. Citado 6 vezes nas páginas 18, 32, 34, 35, 46 e 48.

FENG, Y.; ZHUANG, Y.; PAN, Y. Popular music retrieval by detecting mood. In: *ACM. Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. [S.l.], 2003. p. 375–376. Citado na página 30.

FU, Z.; LU, G.; TING, K. M.; ZHANG, D. A survey of audio-based music classification and annotation. *IEEE transactions on multimedia*, IEEE, v. 13, n. 2, p. 303–319, 2011. Citado na página 28.

GABRIELSSON, A.; LINDSTRÖM, E. The influence of musical structure on emotional expression. Oxford University Press, 2001. Citado na página 28.

GARRIDO, S. A systematic review of the studies measuring mood and emotion in response to music. *Psychomusicology: Music, Mind, and Brain*, Educational Publishing Foundation, v. 24, n. 4, p. 316, 2014. Citado 2 vezes nas páginas 72 e 73.

GIANNAKOPOULOS, T. PyAudioAnalysis: An open-source python library for audio signal analysis. *PLoS ONE*, v. 10, n. 12, p. 1–17, 2015. ISSN 19326203. Citado 2 vezes nas páginas 52 e 70.

GILDA, S.; ZAFAR, H.; SONI, C.; WAGHURDEKAR, K. Smart music player integrating facial emotion recognition and music mood recommendation. In: *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*. [S.l.: s.n.], 2017. p. 154–158. Citado 6 vezes nas páginas 18, 33, 44, 45, 47 e 48.

GONCALVES, V. P. Uma abordagem para indicar o estado emocional de usuários em tempo de interação. In: . [S.l.]: Universidade de São Paulo, 2016. (IHC '12). Citado na página 22.

HSU; CHAWAI, D.; YUNG-JEN, J. {LyQ}: An Emotion-Aware Music Player. *2006 {AAAI} Workshop on Computational Aesthetics: Artificial Intelligence Approaches to Beauty and Happiness*, n. Mehrabian 1995, p. 95–98, 2006. Disponível em: <<http://www.aaai.org/Library/Workshops/ws06-04.php>>. Citado 6 vezes nas páginas 32, 36, 37, 38, 46 e 48.

- IYER, A. V.; PASAD, V.; SANKHE, S. R.; PRAJAPATI, K. Emotion based mood enhancing music recommendation. In: *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*. [S.l.: s.n.], 2017. p. 1573–1577. Citado 6 vezes nas páginas 18, 33, 43, 44, 47 e 48.
- JANSSEN, J. H.; Van Den Broek, E. L.; WESTERINK, J. H. Tune in to your emotions: A robust personalized affective music player. *User Modeling and User-Adapted Interaction*, v. 22, n. 3, p. 255–279, 2012. ISSN 09241868. Citado 2 vezes nas páginas 16 e 17.
- KAMBLE, S. G.; KULKARNI, A. H. Facial expression based music player. *2016 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2016*, p. 561–566, 2016. Citado 7 vezes nas páginas 17, 32, 40, 41, 42, 46 e 48.
- KATAYOSE, H.; IMAI, M.; INOKUCHI, S. Sentiment extraction in music. In: IEEE. *Pattern Recognition, 1988., 9th International Conference on*. [S.l.], 1988. p. 1083–1087. Citado na página 30.
- KIM, Y. E.; SCHMIDT, E. M.; MIGNECO, R.; MORTON, B. G.; RICHARDSON, P.; SCOTT, J.; SPECK, J. A.; TURNBULL, D. Music emotion recognition: A state of the art review. In: *Proc. ISMIR*. [S.l.: s.n.], 2010. p. 255–266. Citado 2 vezes nas páginas 25 e 30.
- LAAR, B. V. D. Emotion detection in music, a survey. In: *Twente Student Conference on IT*. [S.l.: s.n.], 2006. v. 1, p. 700. Citado na página 29.
- LEONARD, M. *Emotion and meaning in music*. [S.l.]: Chicago: University of Chicago Press, 1956. Citado na página 25.
- LI, T.; OGIHARA, M. Detecting emotion in music. Johns Hopkins University, 2003. Citado 2 vezes nas páginas 29 e 30.
- LIN, C.; LIU, M.; HSIUNG, W.; JHANG, J. Music Emotion Recognition Based on Two-Level Support Vector Classification. p. 10–13, 2016. Citado na página 16.
- LIU, H.; SINGH, P. Conceptnet—a practical commonsense reasoning tool-kit. *BT technology journal*, Springer, v. 22, n. 4, p. 211–226, 2004. Citado na página 36.
- LOCKNER, D.; BONNARDEL, N.; BOUCHARD, C.; RIEUF, V. Emotion and interface design. *Proceedings of the 2014 Ergonomie et Informatique Avancée Conference - Design, Ergonomie et IHM: quelle articulation pour la co-conception de l'interaction on - Ergo'IA '14*, p. 33–40, 2014. ISSN 18729126. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2671470.2671475>>. Citado na página 22.
- LU, L.; LIU, D.; ZHANG, H.-J. Automatic mood detection and tracking of music audio signals. *IEEE Transactions on audio, speech, and language processing*, IEEE, v. 14, n. 1, p. 5–18, 2006. Citado 2 vezes nas páginas 17 e 28.
- MACDORMAN, K. F. Stuart ough chin-chang h, “automatic emotion prediction of song excerpts: Index construction, algorithm design, and empirical comparison,”. *Journal of New Music Research*, v. 36, n. 4, 2007. Citado na página 27.
- MAHLKE, S.; MINGE, M. Consideration of multiple components of emotions in human-technology interaction. In: \_\_\_\_\_. *Affect and Emotion in Human-Computer Interaction: From Theory to Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. p. 51–62. ISBN

978-3-540-85099-1. Disponível em: <[https://doi.org/10.1007/978-3-540-85099-1\\_5](https://doi.org/10.1007/978-3-540-85099-1_5)>. Citado na página 30.

MCDUFF, D.; MAHMOUD, A.; MAVADATI, M.; AMR, M.; TURCOT, J.; KALIOUBY, R. e. Affdex sdk: A cross-platform real-time multi-face expression recognition toolkit. In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2016. (CHI EA '16), p. 3723–3726. ISBN 978-1-4503-4082-3. Disponível em: <<http://doi.acm.org/10.1145/2851581.2890247>>. Citado na página 71.

MCKEOWN-GREEN, J. What is music? is there a definitive answer?. *Journal of Aesthetics & Art Criticism*, v. 72, n. 4, p. 393 – 403, 2014. ISSN 00218529. Citado na página 16.

MILETTO, E. M.; COSTALONGA, L. L.; FLORES, L. V.; FRITSCH, E. F.; VICARI, M. S. P. R. M. Introdução à Computação Musical. v. 2004, 2004. Citado na página 24.

NIELSEN, J. *Quantitative Studies: How Many Users to Test?* 2006. Disponível em: <<https://www.nngroup.com/articles/quantitative-studies-how-many-users/>>. Citado na página 67.

NORMAN, D. *Emotional Design: Why We Love (or Hate) Everyday Things*. [S.l.: s.n.], 2004. v. 27. ISBN 0465051359. Citado na página 22.

NORMAN, D. A. *Emotional Design: Why We Love (or Hate) Everyday Things*. New York, NY: Basic Books, 2005. 257 p. ISBN 9780465051366. Citado na página 21.

OLIVEIRA, T. Modesto Veludo de. *Amostragem não Probabilística: Adequação de Situações para uso e Limitações de amostras por Conveniência, Julgamento e Quotas*. 2001. Disponível em: <[http://www.fecap.br/adm\\_online/art23/tania2.htm](http://www.fecap.br/adm_online/art23/tania2.htm)>. Citado na página 67.

ORTONY, A.; CLORE, G.; COLLINS, A. *The Cognitive Structure of Emotion*. [S.l.: s.n.], 1988. v. 18. Citado na página 22.

PICHL, M.; ZANGERLE, E.; SPECHT, G. Understanding playlist creation on music streaming platforms. *Proceedings - 2016 IEEE International Symposium on Multimedia, ISM 2016*, p. 475–480, 2017. Citado na página 50.

Emotion: Theory, research, and experience. In: PLUTCHIK, R.; KELLERMAN, H. (Ed.). *Theories of Emotion*. Academic Press, 1980. p. ii. ISBN 978-0-12-558701-3. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B9780125587013500016>>. Citado na página 22.

REIJNEVELD, K.; LOOZE, M. de; KRAUSE, F.; DESMET, P. Measuring the Emotions Elicited by Office Chairs. In: *Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces*. New York, NY, USA: ACM, 2003. (DPPI '03), p. 6–10. ISBN 1-58113-652-8. Disponível em: <<http://doi.acm.org/10.1145/782896.782899>>. Citado na página 30.

RENTFROW, P. J.; GOSLING, S. D. The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, v. 84, n. 6, p. 1236–1256, 2003. ISSN 1939-1315. Disponível em: <<http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.84.6.1236>>. Citado na página 17.

RUSSELL, J. A circumplex model of affect. v. 39, p. 1161–1178, 12 1980. Citado 2 vezes nas páginas 22 e 23.

SANTINI, R. M.; SOUZA, R. F. D. RECUPERAÇÃO DA INFORMAÇÃO DE MÚSICA E A CIENCIA DA INFORMAÇÃO: Tendências e desafios de pesquisa. *VIII ENANCIB - Encontro Nacional de Pe*, 2007. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT2--143.pdf>>. Citado na página 17.

SCHERER, K. R. Introduction to Social Psychology. *A European perspective*, v. 44, p. 151–191, 2000. Citado na página 23.

SCHERER, K. R. Appraisal processes in emotion: theory, method, research. *Appraisal considered as a process of multilevel sequential checking*, v. 4, n. 4, p. 92–120, 2001. Citado na página 21.

SCHERER, K. R. What are emotions? And how can they be measured? *Social Science Information*, v. 44, n. 4, p. 695–729, dez. 2005. ISSN 0539-0184. Citado 6 vezes nas páginas 19, 22, 23, 24, 30 e 49.

SCHERER, K. R. Emotions are emergent processes: they require a dynamic computational architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2009. Citado na página 21.

SOLEYMANI, M.; CARO, M. N.; SCHMIDT, E. M.; SHA, C.-Y.; YANG, Y.-H. 1000 songs for emotional analysis of music. In: ACM. *Proceedings of the 2nd ACM international workshop on Crowdsourcing for multimedia*. [S.l.], 2013. p. 1–6. Citado 3 vezes nas páginas 26, 53 e 70.

SOMEREN, M. W. van; BARNARD, Y. F.; SANDBERG, J. A. C. *The Think Aloud Method: A practical guide to modelling cognitive processes*. London, UK: Academic Press, 1994. 218 p. ISBN 0127142703. Citado na página 30.

THAYER, R. E.; MCNALLY, R. J. The biopsychology of mood and arousal. *Cognitive and Behavioral Neurology*, LWW, v. 5, n. 1, p. 65, 1992. Citado na página 28.

WANG, G. The characteristics of network music and its evaluation. *Journal of Zhengzhou University: Philosophy and Social Sciences*, v. 44, p. 73–75, 2007. Citado 2 vezes nas páginas 16 e 50.

WANG, J.-C.; YANG, Y.-H.; WANG, H.-M.; JENG, S.-K. The acoustic emotion gaussians model for emotion-based music annotation and retrieval. In: ACM. *Proceedings of the 20th ACM international conference on Multimedia*. [S.l.], 2012. p. 89–98. Citado na página 27.

XIAO, Z.; DELLANDREA, E.; DOU, W.; CHEN, L. What is the best segment duration for music mood analysis? In: IEEE. *Content-Based Multimedia Indexing, 2008. CBMI 2008. International Workshop on*. [S.l.], 2008. p. 17–24. Citado na página 28.

XUE, H.; XUE, L.; SU, F. Multimodal music mood classification by fusion of audio and lyrics. In: SPRINGER. *International Conference on Multimedia Modeling*. [S.l.], 2015. p. 26–37. Citado na página 29.

YANG, D.; LEE, W.-S. Disambiguating music emotion using software agents. In: *ISMIR*. [S.l.: s.n.], 2004. v. 4, p. 218–223. Citado na página 29.

YANG, T.; HE, C. From Online to Offline: Charactering User's Online Music Listening Behavior for Efficacious Offline Radio Program Arrangement. 2017. Citado na página 50.

YANG, X.; DONG, Y.; LI, J. Review of data features-based music emotion recognition methods. *Multimedia Systems*, Springer, p. 1–25, 2017. Citado 6 vezes nas páginas 17, 25, 26, 27, 29 e 30.

YANG, Y.-H.; SU, Y.-F.; LIN, Y.-C.; CHEN, H. H. Music emotion recognition: The role of individuality. In: ACM. *Proceedings of the international workshop on Human-centered multimedia*. [S.l.], 2007. p. 13–22. Citado 2 vezes nas páginas 25 e 27.

ZENTNER, M.; GRANDJEAN, D.; SCHERER, K. R. Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*, American Psychological Association, v. 8, n. 4, p. 494, 2008. Citado na página 17.

ZHANG, J. L.; HUANG, X. L.; YANG, L. F.; XU, Y.; SUN, S. T. Feature selection and feature learning in arousal dimension of music emotion by using shrinkage methods. *Multimedia Systems*, Springer, v. 23, n. 2, p. 251–264, 2017. Citado 2 vezes nas páginas 27 e 28.

# Apêndice A

## TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

---

---

1. Você está sendo convidado para participar da pesquisa “Um algoritmo para recomendação musical baseado na emoção desejada do usuário”.
2. Você foi selecionado para ser voluntário e sua participação não é obrigatória.
3. A qualquer momento você pode desistir de participar e retirar seu consentimento.
4. Sua recusa não trará nenhum prejuízo em sua relação com o pesquisador ou com a instituição.
5. Esta pesquisa tem por objetivo avaliar o estado emocional do usuário por meio de uma seleção músicas fornecidas pelo próprio usuário. Para realizar a avaliação é necessário que seja coletado as seguintes informações: gravação de voz, expressões faciais e preenchimento de questionário subjetivo
6. A sua participação na pesquisa consistirá em ouvir um conjunto de músicas selecionadas a partir de um domínio de 20 músicas fornecidas a priori. A playlist gerada para a realização desse experimento será limitada entre 5 a 8 minutos.
7. A sua participação na pesquisa poderá envolver os seguintes riscos: estresse, tristeza, cansaço e tédio, seja pelos conteúdos, tempo de realização das tarefas e/ou resposta de questionários. Pode ocorrer também frustração ou impacto negativo pelos itens citados, assim como desconforto por estar sendo avaliado. Será feito o possível para minimizar os impactos listados previamente, desde o planejamento e a seleção dos conteúdos. Ainda assim, caso ocorram, você pode se recusar a responder aos questionários ou mesmo interromper a sua participação a qualquer momento.
8. A aplicação da pesquisa será realizada pelo pesquisador responsável com o apoio de colaboradores do Laboratório de Interação Flexível e Sustentável (LIFeS) da Universidade Federal de São Carlos.

9. Os benefícios relacionados à sua participação estão em contribuir com a pesquisa, como descrito no item 5 desse termo. Você terá acesso aos resultados da pesquisa por meio de artigos científicos publicados em congressos e pelo site do laboratório LIFeS: <http://lifes.dc.ufscar.br>.
10. As informações obtidas através dessa pesquisa serão confidenciais e assegura-se o sigilo sobre sua participação.
11. Os dados não serão divulgados de forma a possibilitar sua identificação.
12. Você receberá uma cópia deste termo em que consta o telefone e o endereço do pesquisador principal, podendo tirar suas dúvidas sobre o projeto e sua participação, agora ou a qualquer momento

---

Willian Garcias de Assunção

Departamento de Computação (DC), Universidade Federal de São Carlos (UFSCar)

Caixa Postal 676, 13565-905, São Carlos, SP – Tel.: (16) 3351 8626

Endereço e telefone do pesquisador principal:

Rua Doutor Marino Costa Terra, 1000 – 13567-640– Parque Sabará, São Carlos, SP

Tel.: (16) 9 9139 5021

Declaro que entendi os objetivos, riscos e benefícios de minha participação na pesquisa e concordo em participar.

O pesquisador me informou que o projeto foi aprovado pelo Comitê de Ética em Pesquisa em Seres Humanos da UFSCar que funciona na Pró-Reitoria de Pós-Graduação e Pesquisa da Universidade Federal de São Carlos, localizada na Rodovia Washington Luiz, Km. 235 - Caixa Postal 676 - CEP 13.565- 905 - São Carlos - SP – Brasil. Fone (16) 3351-8110. Endereço eletrônico: [cephumanos@power.ufscar.br](mailto:cephumanos@power.ufscar.br).

São Carlos, \_\_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_.

---

Assinatura do sujeito da pesquisa ou responsável

# Apêndice B

## AUTORIZAÇÃO DE CAPTAÇÃO E EXIBIÇÃO DE IMAGEM, SOM E NOME

---

---

Eu, \_\_\_\_\_, (nacionalidade) \_\_\_\_\_, (estado civil) \_\_\_\_\_, portador da Cédula de Identidade RG \_\_\_\_\_ (ou Registro Nacional de Estrangeiro \_\_\_\_\_), inscrito no CPF/MF sob o número \_\_\_\_\_, autorizo a captação, utilização e exibição de minha voz e imagem pela UNIVERSIDADE FEDERAL DE SÃO CARLOS – UFSCar, diretamente ou através do Departamento de Computação ou outra entidade vinculada ou contratada, a serem utilizadas em obras audiovisuais a serem produzidas para fins institucionais, didáticos e/ou científicos, sejam essas destinadas à divulgação ao público em geral e/ou apenas para uso interno desta instituição.

1. A presente autorização, concedida a título gratuito, confere à UFSCar, diretamente ou através do Laboratório de Interação Flexível e Sustentável – Departamento de Computação ou outra entidade vinculada ou contratada, o direito de utilizar minhas imagens e voz, nas obras para veiculação interna na UFSCar, bem como em eventos externos, no Brasil e no exterior, por mídia escrita, eletrônica ou digital, tais como Revistas, Manuais, Portais de Internet, folders, atividades de caráter didático ou científico, trabalhos científicos, publicações em geral, entre outros, a critério exclusivo da UFSCar, desde que não haja desvirtuamento de sua finalidade.
2. As obras poderão ser distribuídas pelo Laboratório de Interação Flexível e Sustentável – Departamento de Computação – UFSCar ou por qualquer outra entidade vinculada ou contratada, de forma gratuita ou comercial, sendo certo que nada será devido pelo uso das imagens ou voz, objetos da presente autorização, mesmo nas hipóteses de comercialização das obras, em parceria ou não com outras pessoas jurídicas.
3. Declaro estar ciente de que as imagens e voz captadas nesta ocasião farão parte de um “banco de imagens” pertencente ao Laboratório de Interação Flexível e Sustentável – Departamento de Computação – UFSCar e que poderão ser utilizadas a qualquer tempo e de acordo com os critérios da mesma.

4. Por esta ser a expressão de minha vontade, declaro que autorizo o uso acima descrito sem que nada haja a ser reclamado a título de direitos conexos à minha imagem ou a qualquer outro título, e assino a presente autorização em duas vias de igual teor e forma.

São Carlos, \_\_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_.

\_\_\_\_\_  
(Assinatura)

\_\_\_\_\_  
(Nome por extenso)

# Apêndice C

## ESPAÇO EMOCIONAL SEMÂNTICO DE SCHERER

Participante: \_\_\_\_\_ Data: \_\_/\_\_/2018 Sexo: M ( ) F ( )  
 Idade: \_\_ Escolaridade \_\_\_\_\_ Gêneros musicais: \_\_\_\_\_

