

Abordagem para a interação afetiva: um estudo de caso com player de música

Luana Okino Sawada

*Instituto de Ciências Matemáticas e de
Computação
Universidade de São Paulo - ICMC/USP
Brazil*

Leandro Yukio Mano

*Instituto de Ciências Matemáticas e de
Computação
Universidade de São Paulo - ICMC/USP
Brazil*

Jó Ueyama

*Instituto de Ciências Matemáticas e de
Computação
Universidade de São Paulo - ICMC/USP
Brazil*

Abstract— Estudos atuais na área de Interação Humano-Computador evidenciam a importância de se considerar aspectos emocionais na interação com sistemas computacionais. Acredita-se que ao permitir agentes artificiais identificar emoções de usuários torna-se possível induzir e despertar emoções a fim de estimulá-los em suas atividades. Um dos grandes desafios dos pesquisadores é prover sistemas capazes de reconhecer, interpretar e reagir de modo inteligente e sensível às emoções do usuário. Nesse sentido, propomos um player de música para a interação afetiva com o usuário baseado no reconhecimento emocional por meio da classificação das expressões faciais. Logo, este projeto tem como objetivo desenvolver e avaliar um sistema que possa identificar o estado emocional do usuário e prover um mecanismo de persuasão com vistas a mudar o estado emocional do usuário (com um estudo de caso em player de música). Ainda, explorar a abordagem flexível na persuasão por meio de mecanismos persuasivos que poderão variar entre um player de música, jogos e/ou vídeos. Assim, ao longo do estudo, o modelo baseado em Comitê de Classificação se mostrou eficiente na identificação das emoções básicas e na satisfação dos usuários mediante a aplicação com o player de música.

Abstract— Current studies in the field of Human-Computer Interaction highlight the relevance of emotional aspects while interacting with computers systems. It is believed that by allowing intelligent agents to identify user's emotions it becomes possible to induce and arouse emotions itself in order to stimulate the users in their activities. One of the great challenges for researchers is to provide systems capable of recognizing, interpreting and reacting intelligently and sensitively to the emotions of the user. In this sense, we propose an affective interaction between the music player and the user that is based on emotional recognition through the interpretation of facial expressions. Therefore, this project aims to develop and evaluate a system that can identify the user's emotional state and provide a persuasive mechanism to change it (with a case study in music player). Also, explore the flexible approach in persuasion through persuasive mechanisms that may vary between a music player, games and / or videos. Thus, throughout the study, the model based on the Classification Committee was efficient in identifying the basic emotions and the satisfaction of the users through the application with the music player.

Keywords—*Emotion Classification, Human-Computer Interaction (HCI), FaceTracker, Music Player.*

I. INTRODUÇÃO

A Emoção é um aspecto fundamental do ser humano, que o permite desfrutar de experiências, além de influenciar suas decisões diárias. Esse aspecto é importante na interação entre as pessoas [1] e pode ser utilizado para intervir na realização de uma tarefa, ou até mesmo para tomar decisões em prol do usuário; por exemplo, sugerindo um gênero de filme para aliviar o estresse com base na sua atual emoção [2].

Os aspectos emocionais vêm sendo estudados há muito tempo na área da Psicologia [3]. No entanto, somente nos últimos anos, houve um aumento significativo de pesquisas sobre o tema na área de Computação Afetiva [1], [4], [5], [6], [7]. Uma das divisões básicas da Computação Afetiva é a Interação Humano-Computador (IHC) [8], a área que estuda como os dispositivos computacionais podem reconhecer, modelar e responder as emoções humanas (dentre outros aspectos) e expressá-las mediante uma interface/interação computacional.

Diante desse panorama, estudos atuais na área de IHC evidenciam a importância de se considerar aspectos emocionais na interação com sistemas computacionais [9], [10]. Analisada por Pesquisadores da *University of Southern California*, na matéria publicada pelo *planettech*, a face desenvolve características diferentes de acordo com estados emocionais distintos. Tais características podem ser analisadas por sensores específicos, de maneira a atribuir uma emoção de acordo com a face do usuário. Esses sistemas computacionais podem ser introduzidos na descrição comportamental e emocional do usuário, utilizando-se desses dados para a melhor satisfação pessoal do indivíduo na interação com dispositivos computacionais.

O principal objetivo de se promover essa interação afetiva é contribuir para o aumento da coerência, da consistência e da credibilidade das reações e das respostas computacionais providas durante a interação humana, via interface humano-computador. Por exemplo, se o computador “percebe” que o usuário está cansado, estressado ou triste, esse tenta “persuadi-lo” à mudança de estado a partir de um player de música, jogos, chat com amigos, filmes e/ou redes sociais.

Assim, sistemas capazes de reconhecer as emoções do usuário, interpretando-as e reagindo de modo inteligente e sensível, poderiam atender mais adequadamente aos requisitos de interação dos indivíduos. Ao permitir que agentes artificiais compreendam emoções em uma interação humano-computador, é possível “persuadir” (induzir e/ou despertar) emoções em usuários de sistemas computacionais [11]. O conhecimento sobre a persuasão permite identificar emoções provocadas no uso de sistemas interativos e pode ser aplicado na concepção e no desenvolvimento de tais sistemas, fomentando, assim, mudanças de atitude e de comportamento nos indivíduos [12].

As emoções são complexas, difíceis de identificar e de mensurar; estão ligadas à reação de vários componentes que podem ser sintetizados em: avaliações cognitivas, sentimentos subjetivos, tendências comportamentais, reações fisiológicas e expressões motoras [13], [14] e trazem à tona a discussão sobre o que as emoções são e o que precisa ser estudado quando se quer determinar o significado das mesmas [13]. De maneira mais específica, o uso de imagens da face dos usuários para

identificar as emoções permite uma interação pervasiva. Isso porque tal informação pode ser obtida sem uma ação explícita do usuário e sem o uso de equipamentos específicos ou intrusivos. Além disso, no âmbito tecnológico, é comum que equipamentos computacionais (por exemplo, tablet, desktop, notebook e smartphone) possuam câmeras frontais. Tipicamente, esses dispositivos permanecem direcionados à face do usuário, podendo ser utilizados para a aquisição de imagens em tempo real.

Nesse sentido, esse artigo explora o desenvolvimento de sistemas que permitem alterações em seu comportamento durante a interação em tempo de execução e que oferecem a cada usuário a possibilidade de ter o sistema ajustado de acordo com suas preferências, necessidades e intenções de uso. Entretanto, ainda é um grande desafio medir o quão persuasiva uma solução é e o quanto ela pode influenciar na tomada de decisão ou na mudança de comportamento dos indivíduos. Assim, este trabalho tem como objetivo desenvolver e avaliar um sistema que possa: **i.)** identificar o estado emocional do usuário; **ii.)** prover um mecanismo de persuasão que objetiva mudar o estado emocional do usuário (com um estudo de caso em player de música) e **iii.)** explorar a abordagem flexível na persuasão (de acordo com o estado emocional particular de cada usuário), através de mecanismos persuasivos que poderão variar entre um player de música, jogos e/ou vídeos.

Este artigo está organizado da seguinte forma: na Seção II destacamos os trabalhos relacionados ao contexto deste artigo. A Seção III aborda conceitos prévios importantes para o entendimento do uso da face para o artigo proposto. Após a apresentação desses conceitos, na Seção IV propomos o sistema (player de música) para interação afetiva por meio da emoção identificada. Já na Seção V, apresentamos a metodologia e os resultados da interação do usuário com o sistema. Por fim, na Seção VI, descrevemos as conclusões e alguns trabalhos que poderão ser desenvolvidos futuramente.

II. TRABALHOS RELACIONADOS

Pesquisas na área de *Machine Learning* (ML) têm gerado uma ampla quantidade de métodos capazes de extrair padrões a partir de exemplos passados e que podem ser utilizados para tomar decisões em situações novas (ainda não conhecidas). As técnicas utilizadas para problemas de classificação estão cada vez mais sendo aplicadas para a análise de respostas emocionais [1], [4], [5], [6], [7], [15].

Um exemplo disso é a utilização de sensores fisiológicos para a detecção e classificação da emoção do usuário. [6] apresentam uma abordagem geral de reconhecimento de emoções utilizando tais sensores. Utilizando essa abordagem, os autores desenvolveram o EREC (Sistema de Reconhecimento de Emoções). Esse sistema coleta dados de alguns tipos de sensores: atividade eletrodérmica, temperatura do ar atmosférico e variabilidade da frequência cardíaca. [6] trazem duas aplicações utilizando o sistema EREC: uma ferramenta de avaliação de usabilidade emocional e um ambiente de *e-learning* emocional. Similarmente, [1] fazem um estudo com o objetivo de prever diferentes estados emocionais do usuário por meio de reações fisiológicas comparando três tipos de modelos

computacionais. Um experimento foi desenvolvido para provocar respostas emocionais com fotos de emoções padronizadas quando algumas reações fisiológicas foram medidas. Três algoritmos de classificação – Regras de Decisão, *k-Nearest Neighbor* (kNN) e Árvore de Decomposição – foram aplicados para a construção de modelos de previsão das características fisiológicas extraídas.

Outra abordagem utilizada é a detecção e classificação da emoção por meio da tendência comportamental do usuário. [4] sugerem uma solução computacional para auxiliar os usuários a gerenciar e mudar o estado emocional da frustração. O sistema solicita informações a todo instante sobre o estado emocional do indivíduo durante a interação e fornece *feedback* ao usuário, especialmente sobre seu conteúdo emocional. O objetivo é incentivar o usuário a sentir-se satisfeito e informar que ele não é o único a ter esse tipo de emoção. [4] trazem em sua proposta uma abordagem que percebe o estado emocional do usuário por meio de mensagens fornecidas durante a interação. Similarmente, [7] utilizam um modelo de reconhecimento de humor baseado em seis tipos de informações: SMS, e-mail, registro de ligações, uso de aplicações, navegação na web e localização. Segundo [7] a detecção automática de humor permite a criação de novos aplicativos e a melhoria de outros existentes. Um aplicativo que registra o humor do usuário e permite que seja consultado pelo próprio usuário em outras situações, por exemplo: local e tempo, pode melhorar a consciência do usuário sobre suas mudanças de humor. O MoodScope se diferencia de outras aplicações de detecção de humor por não utilizar dados de sensores físicos, como os dos smartphones, e os dados de áudio e vídeo.

A pesquisa proposta por [5] propõe e apresenta um modelo baseado em algoritmos de classificação – Regressão Linear e Rede Neural Artificial – para análise em tempo real de emoções baseando-se em expressões motoras e reações fisiológicas. A análise foi feita a partir de vídeos de 41 indivíduos assistindo filmes altamente emotivos. Um dos sensores faz o reconhecimento de emoções faciais e outro faz análise da variabilidade da frequência cardíaca e da atividade eletrodérmica do usuário. Cinco especialistas fizeram a análise dos vídeos com o objetivo de definir as classes (saídas) para os algoritmos de classificação, por exemplo, analisando se o usuário estava triste em determinado momento. [5] apresentam algoritmos baseados em pontos extraídos do rosto e respostas fisiológicas do usuário. Similarmente, [15] investigam o reconhecimento de emoções com base em conjuntos de elementos da expressão facial. [15] mostram que, para as características, tais como as formas dos componentes faciais (olhos, boca, nariz, queixo e sobrancelhas) e a localização dos pontos salientes faciais (cantos dos olhos e boca) foram utilizados para propor seis representações faciais distintas, que enfatizam alterações faciais e são baseadas em estudos psicológicos. [15] evidenciam que o uso de áreas pré-definidas da face em conjunto com ângulos e distâncias é uma proposta válida para construir modelos de classificação da emoção.

Diante do exposto nessa seção, observou-se que cada vez mais surgem estudos que contribuem com técnicas para o desenvolvimento de sistemas capazes de identificar, por meio de sensores e/ou informações, uma determinada emoção do

usuário de dispositivos computacionais. Além disso, tais estudos também se relacionam pela classificação da emoção ao permitirem que decisões sejam tomadas com maior precisão diante a interação do usuário e os dispositivos computacionais. Deve-se ressaltar, entretanto, a ausência de trabalhos que utilizam essa abordagem para a interação homem-máquina de maneira que sistemas computacionais possam interagir com o usuário de acordo com a emoção detectada. Dessa forma, com o objetivo de abordar essa lacuna e com o intuito de realizar essa interação afetiva e de forma ubíqua, propomos um sistema que identifica a emoção do usuário pela face e realiza a interação para melhor satisfazer as intenções do usuário.

III. CONHECIMENTO SOBRE EMOÇÕES

A emoção é uma reação complexa de um indivíduo relacionada diretamente com as suas necessidades de bem-estar e envolve todas as partes do corpo desse ser. Assim, a fim de determinar o significado das emoções, diversos componentes são estudados em pesquisa sobre esse tema [13], [16]. Alguns dos subtemas mais estudados são: antecedentes das emoções e avaliação de situações [17], [18], [19]; termos do bem-estar de um indivíduo [20]; padrões de reação psicofisiológica, ou seja, estudo das relações entre fenômenos psíquicos e fisiológicos [21], [22]; e experiências subjetivas [23], como por exemplo, inibição e no controle sobre as mesmas [16].

De acordo com a Teoria Componencial das Emoções [13], [22], [24], as emoções são como diferentes níveis de mudanças inter-relacionadas entre um conjunto de cinco componentes [14], [19]. São esses:

- 1) *Avaliação Cognitiva*: é ligada à interpretação de uma situação. O indivíduo avalia constantemente o mundo ao seu redor e procura perceber as qualidades afetivas de eventos, objetos e recursos;
- 2) *Experiência Subjetiva*: é responsável pela avaliação de um episódio emocional, ou seja, da experiência consciente que se relaciona com a capacidade do indivíduo de regular as emoções;
- 3) *Tendências Comportamentais*: preparam e orientam o indivíduo para (ou durante) uma determinada ação, como, por exemplo, o tempo necessário para completar uma tarefa, o número de erros cometidos, entre outros aspectos;
- 4) *Reações Fisiológicas*: são aspectos emocionais que permitem ao usuário revelar espontaneamente e inconscientemente suas emoções.
- 5) *Expressões Motoras, também conhecidas como Reações Expressivas*: implica em alterações nas expressões faciais, vocais e/ou gestuais que acompanham a experiência emocional do usuário. Assim, o rosto e a voz sofrem mudanças com o grau de excitação.

Uma abordagem baseada em componentes oferece uma maneira mais abrangente para entender os diferentes aspectos das emoções em todos os tipos de situações [14]. Assim, diferentes aspectos podem ser utilizados para a análise emocional sem, no entanto, desconsiderar a relação entre esses diferentes componentes [19].

Nesse sentido, devido a facilidade da captura das imagens e a oportunidade na aplicação de Computação Ubíqua na interação dos usuários por dispositivos computacionais (tablet, desktop, notebook, smartphone e sistemas embarcados), os quais normalmente possuem uma câmera e a interação com os usuários são realizadas em frente ao dispositivos, neste trabalho foi considerado características da face do usuário, ou seja, aspectos emocionais relacionados as **Expressões Motoras**.

Como dito anteriormente, as emoções são complexas, difíceis de identificar e de mensurar. A fim de classificá-las, o reconhecimento da emoção é baseado na ciência psicológica e cognitiva, que afirma haver duas visões primárias sobre a representação das emoções: i.) contínuas e ii.) categóricas [25].

A visão contínua aborda as emoções como pontos em um espaço multidimensional e utiliza escalas contínuas ou bases dimensionais. Nesse ponto de vista, estados afetivos são sistematicamente relacionados entre si, não discretos e independentes uns dos outros. A abordagem dimensional mais comum é o modelo *circumplex* [25] (Figura 1), o qual argumenta que todas as emoções se encontram em um espaço bidimensional contínuo e suas dimensões são valência (positiva ou negativa, eixo X) e excitação (o nível de energia ou de excitação associado à emoção, eixo Y). Resumidamente, o modelo é composto por uma estrutura em formato de círculo, por meio de um escalonamento em dois domínios principais: prazer e desprazer.



Fig. 1. Modelo *circumplex* proposto por [25].

Na representação categórica, diferentes emoções são mapeadas em categorias distintas que utilizam os componentes da Teoria Componencial das Emoções. Neste trabalho foram utilizadas as *Expressões Motoras*, o componente responsável por comunicar as tendências comportamentais do indivíduo [11], [13], [14]. Esse componente das emoções implica em alterações nas expressões faciais que acompanham a experiência emocional do usuário. O exemplo mais comum dessa descrição são as seis emoções básicas (alegria, aversão, medo,

raiva, surpresa e tristeza) e o conjunto de expressões faciais relacionadas a elas, sendo inatas e universais entre as culturas. O termo *universal* indica que os mesmos movimentos musculares são produzidos ao realizar as expressões faciais. Todas as outras categorias emocionais são, então, construídas a partir de combinações dessas emoções básicas. Essa abordagem é apoiada pelos estudos transculturais, os quais indicam que os seres humanos percebem certas emoções básicas transmitidas por expressões faciais da mesma forma, independentemente da cultura, da etnia e dos costumes [26].

As emoções são identificadas por seres humanos de maneiras distintas e os autores classificam as emoções em três grupos: (i) alegria e surpresa, fáceis de identificar; (ii) raiva e tristeza, um pouco mais difíceis e (iii) medo e aversão, ainda mais difíceis [27]. A seguir, na Figura 2, encontram-se as emoções analisadas. O estado neutro também é considerado, o qual é utilizado como referência para a detecção de estados emocionais.

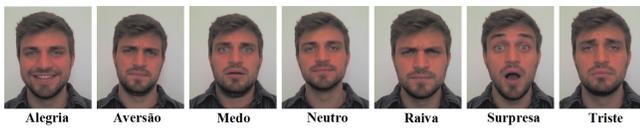


Fig. 2. Emoções básicas [28].

É importante notar que ambas as representações de emoções têm sido utilizadas eficientemente em várias aplicações computacionais [2], [13], [15], [29]. De fato, procedimentos de classificação têm ajudado na análise de repostas emocionais, auxiliando médicos no diagnóstico da depressão, usuários que buscam uma mudança de comportamento em sua rotina, entre outros.

IV. PROPOSTA PARA INTERAÇÃO DE SISTEMAS COMPUTACIONAIS

Aliado ao crescimento dos usuários de sistemas computacionais e ao acesso facilitado aos meios de telecomunicação, têm-se a possibilidade de capturar e de analisar dados de indivíduos em diversos contextos. Diante disso, a detecção da emoção pode ser utilizada para adaptar o comportamento de sistemas computacionais as necessidades de cada usuário ou “persuadí-lo” à transição para uma emoção positiva (por exemplo, de neutro para feliz). Sendo uma proposta que visa a maior interação pessoal com equipamentos tecnológicos, fica evidente a inovação dos métodos de interação do usuário, dentre os vários softwares de uso diário.

Assim, a detecção da emoção pode ser utilizada para o desenvolvimento de novas maneiras de interação do usuário e, posteriormente, caminhar para os dispositivos cognitivos [30]. Tecnologias da informação que possam intervir, caso o usuário esteja estressado ou mau humorado, indicando músicas que possam acalmá-lo, por exemplo e, assim, deixá-lo um pouco mais feliz e/ou reduzir o estresse é um dos caminhos para atender as preferências e as necessidades dos usuários.

A. Reconhecimento da Face

Neste trabalho, a abordagem para a identificação do rosto e para a análise da expressão facial é baseada em características geométricas. Assim, é possível modelar a forma e a localização de componentes faciais pelo uso de pontos característicos e de elementos geométricos, como ângulos, distâncias ou áreas utilizadas para a representação geométrica do rosto.

Os métodos baseados em características geométricas são utilizados na modelagem facial (expressões motoras) com o objetivo de desenvolver uma abordagem que se assemelhe à maneira como os seres humanos interpretam os elementos da face. Assim, diferentes representações faciais (alegria, aversão, medo, neutro, raiva, surpresa e tristeza) são propostas para a identificação e a classificação das emoções, codificando a configuração facial de um indivíduo.

O *FaceTracker* [31] é caracterizado como um sistema de visão computacional utilizado para obter as informações faciais. Ele faz uso de uma estratégia de otimização que emprega aproximação linear e ajusta pontos de referência em localizações consistentes para registrar um modelo parametrizado. Com base em um modelo de referência facial composto por 66 pontos característicos, o algoritmo busca alinhar os elementos da face em análise aos pontos característicos do modelo de referência. Nesse sentido, o algoritmo do *FaceTracker* foi modificado de modo a mapear somente um subconjunto de 33 dos 66 pontos faciais obtidos inicialmente [15], [32], o que diminui o custo computacional com a eliminação de possíveis redundâncias. A Figura 3 apresenta um exemplo do mapeamento da face realizada pelo algoritmo modificado do *FaceTracker*.

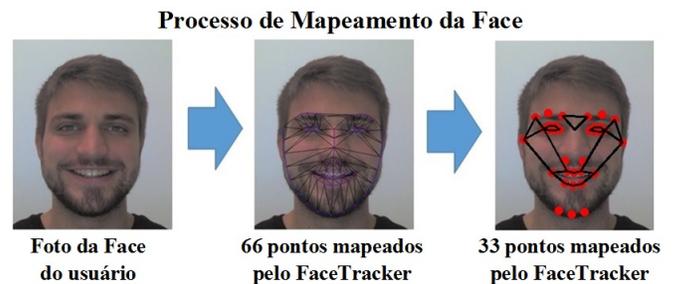


Fig. 3. Processo de mapeamento da face pelo *FaceTracker* [32].

Os elementos considerados na representação facial têm como objetivo modelar as partes do rosto que são essenciais na relação dos movimentos e a emoção expressada de acordo com psicólogos [28]. Portanto, o mapeamento da face utiliza oito pontos para a boca, seis para cada um dos olhos, três para cada sobrancelha e o queixo, dois para as narinas e dois pontos delimitando duas extremidades laterais da face perto dos olhos. Destaca-se que todas as combinações possíveis de pontos são obtidas por meio dos ângulos e das distâncias entre dois pontos distintos, o que gera uma representação com dimensionalidade $D_1 = 2 \times 33 + 8 + 2 \times 528 = 1130$.

B. Classificação da Emoção

A área de classificação de dados visa construir sistemas que aprendam com os dados. No contexto de reconhecimento de emoções, os algoritmos de classificação utilizam o tipo de aprendizado supervisionado, os quais são treinados com exemplos rotulados, isto é, a saída da entrada é conhecida, por exemplo, uma imagem que contém uma face com expressão já conhecida [33], [34]. Para a classificação de emoções diversas são as técnicas de classificação aplicadas à análise de respostas emocionais, sendo elas: *kNN*, *Árvore de Decisão*, *Lógica Fuzzy*, *Redes Bayesianas e SVM* [1], [2], [35], [36], [37]. No entanto, alguns trabalhos abordam o conceito de Comitê de Classificação, do Inglês *Ensemble of Classification* (EC) [29], [32], [38]. Esse fato se deve à premissa de que a combinação de classificadores pode conduzir a uma melhoria no desempenho em termos de melhor generalização e/ou de aumento da acurácia [39].

Um dos aspectos que garante um bom desempenho de um EC é a diversidade entre seus componentes [40]. Para que um Comitê obtenha um desempenho aceitável, ele deverá ser formado por classificadores que possuam acurácia razoável e que não cometam erros coincidentes, para que os erros de um classificador possam ser corrigidos pela opinião de todos os outros componentes [38], [39], [40]. A Figura 4 ilustra o módulo baseado em EC.

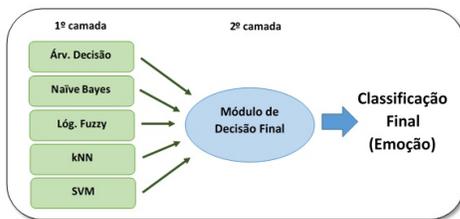


Fig. 4. Módulo baseado em Comitê de Classificação para a classificação da emoção da face do usuário [32].

Tendo como entrada o resultado do mapeamento facial realizado pelo *FaceTracker*, a primeira camada de processamento consiste em classificadores individuais que podem ter diferentes arquiteturas, mas que apresentam saídas de forma comum, isto é, como a indicação de uma classe (emoção) a partir do viés de classificação do respectivo algoritmo. A segunda camada de processamento consiste em uma tomada de decisão que opera nos resultados da camada anterior para gerar a decisão geral do EC. Vale salientar que, para a decisão final do Comitê, foi utilizada a Equação 1 no processo de ponderação e foram atribuídos pesos proporcionais a cada algoritmo de acordo com a sua taxa de acurácia. Na equação, \bar{W} é a média ponderada da acurácia de cada algoritmo e AR corresponde à taxa média da acurácia do algoritmo. Assim, para cada algoritmo utilizado, um peso proporcional terá sido atribuído à sua taxa de acurácia.

$$\bar{W}x_i = \frac{ARx_i}{\sum_j^n ARx_j} \quad (1)$$

Para avaliar o avanço e a acurácia obtidos com a utilização

do módulo de classificação baseado na face, implementamos os algoritmos citados, bem como o módulo de decisão do EC. Com isso, coletamos dados sobre o comportamento e a avaliação do módulo de classificação como um todo.

C. Interação Flexível baseado na Emoção

Esse trabalho discute o uso de uma abordagem genérica para a construção de aplicações adaptativas em dispositivos computacionais. A reconfiguração em tempo de execução é uma característica fundamental para lidar com a heterogeneidade de hardware inerente a dispositivos com recursos escassos (smartphone e nós sensores) [41]. Sendo assim, uma camada de software denominada foi desenvolvida baseada no modelo de componentes do *OpenCom* [41], a qual apresenta uma camada de software genérico; uma arquitetura flexível e extensível; independência de linguagem e; é baseada em *microkernel*, onde as funcionalidades são incrementadas quando necessário. Portanto, apoia o desenvolvimento de sistemas que se adaptam ao estado emocional do usuário durante a interação [41]. Na arquitetura o *microkernel* instância e realiza a conexão dos componentes a fim de “persuadir” o usuário. No caso deste trabalho instância e conecta um player de música, mas podendo ser desenvolvidos outros softwares para interação com o usuário, por exemplo jogos (ver Figura 5).

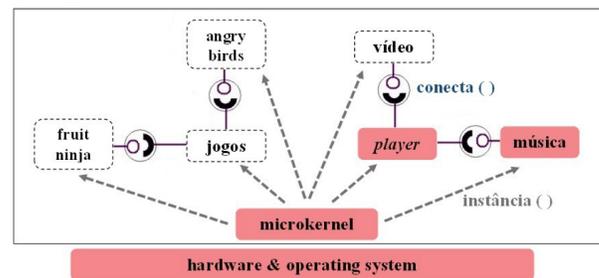


Fig. 5. Cenário de interação de componentes do sistema.

Assim, temos como objetivo possibilitar a interação de softwares, de maneira a “persuadir” o usuário, por exemplo, acesso a redes sociais, chats de bate-papo, jogos ou player de música. No caso do player de música, estudo de caso desta pesquisa, o modelo carrega a playlist (conjunto de músicas) de gênero musical que melhor se adaptar ao estado emocional do usuário naquele momento, de acordo com os dados transmitidos pelo *framework* de classificação. Por exemplo, se o *framework* classifica a experiência emocional do usuário como “triste”, o player irá carregar uma playlist para tentar “persuadí-lo” a baixar o nível de tristeza e até mesmo para uma mudança de estado emocional. Vale salientar que devido a abordagem estudada não se ater a um dispositivo computacional específico, mas com o intuito de abranger o maior número possível de dispositivos (por exemplo, tablet, desktop, notebook e smartphone) o módulo de reconhecimento e classificação da emoção foi implementado em um servidor remoto aplicando conceitos de *Cloud Computing*. Assim, independente da arquitetura e

poder de processamento do dispositivo, o usuário pode fazer uso do sistema.

Diante desse contexto, o sistema interage com o usuário para proporcionar a ele um sentimento de satisfação dentro do contexto em que se encontra. Esta proposta foi validada com um estudo de caso, onde um player de música propõe uma playlist com o objetivo de estimular o usuário de maneira positiva. O modelo desenvolvido considera como entrada um estado emocional do usuário, reconhecido por meio de um *framework* para a classificação das emoções, para a classificação automatizada de emoções através do reconhecimento das características da face do usuário.

Assim, desenvolveu-se uma página Web para os testes de reconhecimento da emoção e para a interação com o player de música. *Ruby on Rails* foi a linguagem de programação escolhida para a implementação devido à experiência com tal linguagem. Seguindo a estrutura proposta, o desenvolvimento da página Web para a interação foi dividido em módulos distintos, descritos a seguir:

- **Preferências do Usuário:** o usuário informa suas preferências musicais para que o software possa disponibilizar as músicas que mais se adaptam ao seu estado emocional, visto que cada usuário possui suas próprias preferências individuais. Esse indica, por meio de um formulário, os artistas que costuma ouvir de acordo com o seu estado emocional. As emoções indicadas no formulário são as tratadas neste trabalho (alegria, aversão, medo, neutro, raiva, surpresa e tristeza).
- **Aquisição da Imagem do Usuário:** é solicitado ao usuário o acesso à câmera de seu dispositivo computacional (móvel ou fixo) e suas imagens são adquiridas em intervalos de 5 segundos. As imagens são enviadas ao servidor, onde são processadas e classificadas, e sua emoção é identificada e armazenada em um banco de dados (para os testes, foi utilizado o banco de dados relacional MySQL). O tempo de 5 segundos foi utilizado de forma empírica, levando em conta somente o tempo para a transmissão e a classificação da imagem. Além disso, o sistema disponibiliza ao usuário a possibilidade de verificar a classificação da sua face em tempo de execução.
- **Player de Música:** já de posse das preferências musicais e da imagem da face do usuário, o player executa as músicas que mais se adaptam ao seu estado emocional. Visto que as imagens são classificadas a cada 5 segundos, as emoções são armazenadas em um arquivo temporário e, ao término de cada música, o sistema verifica qual emoção foi mais detectada nesse intervalo de tempo e a próxima música a ser tocada correspondente as preferências informadas. O player disponibiliza as músicas de duas formas: i) áudio, por meio de um módulo do Spotify® e ii) áudio e vídeo, por meio do Youtube®.
- **Visualização dos Dados:** o sistema ainda disponibiliza um gráfico em tempo de execução que fornece ao usuário um *feedback* das emoções classificadas ao longo do tempo.

V. AVALIAÇÃO E DISCUSSÃO

Visando verificar a proposta de interação de acordo com a emoção do usuário, um experimento comparativo foi executado com um conjunto de 30 indivíduos, com idade entre 18 e 80 anos. No início da avaliação da interação foi realizado um planejamento a fim de verificar a satisfação/bem-estar do usuário diante a interação com o player de música. Vale salientar que esse experimento, junto com um aluno de Doutorado, Vinícius Pereira Gonçalves orientado pelo Prof. Jó Ueyama, foi submetido e autorizado pelo Comitê de Ética em pesquisa envolvendo Seres Humanos da Faculdade de Medicina da Universidade de São Paulo (CAAE-45081415.0.0000.0065).

- **Hipótese:** tendo como base a diversidade e o aumento da utilização de dispositivos computacionais, acredita-se que a proposta é uma solução que contempla situações específicas (emoção), no que diz respeito à interação do usuário com dispositivos computacionais.
- **Objetivo:** observar e analisar a satisfação na interação dos usuários com dispositivos computacionais de acordo com as suas emoções.
- **Metodologia:** com o intuito de analisar a interação do usuário com sistemas flexíveis, um grupo de 30 usuários, com idade entre 18 e 80 anos, foi convidado a participar de uma prática de uso do Notebook. Esses usuários foram divididos em dois grupos. O grupo 1 realizou o experimento com a atuação do sistema flexível na interação com o player de música, enquanto o grupo 2 realizou a interação sem a interferência, ou seja, as músicas foram tocadas de modo totalmente aleatório. A Tabela I apresenta a forma que os dois grupos foram divididos.

TABLE I. DIVISÃO DOS GRUPOS PARA OS EXPERIMENTOS.

	18 a 30 anos	31 a 50 anos	mais de 50 anos
Grupo 1	5	5	5
Grupo 2	5	5	5

Os experimentos foram divididos dessa maneira, de modo a verificar a satisfação do usuário na interação com o player de música. O objetivo da atividade era que esses usuários utilizassem o Notebook conforme a sua preferência, enquanto ouviam músicas no sistema desenvolvido.

- **Material de Apoio:** para a condução do experimento exploratório, foram preparados um Termo de Consentimento Livre e Esclarecido, uma Autorização de Captação e Exibição de Imagem e um Formulário de Observação do Participante. O Termo de Consentimento Livre e Esclarecido elucida aos participantes o objetivo da pesquisa, a voluntariedade na participação e o caráter científico da mesma. Autorização de Captação e Exibição de Imagem concede a exploração das imagens adquiridas realizadas no decorrer dos testes. O Formulário de Observação do Participante foi elaborado para auxiliar no experimento exploratório durante a interação do usuário com o dispositivo computacional.
- **Dispositivo utilizado:** os usuários foram organizados de maneira que realizassem o experimento de forma

TABLE II. MATRIZ DE CONFUSÃO REFERENTE À CLASSIFICAÇÃO DO COMITÊ DAS IMAGENS DOS USUÁRIOS SUBMETIDOS AO TESTE.

	A	B	C	D	E	F	G
Alegria = A	30	0	0	0	0	0	0
Aversão = B	2	21	0	0	4	0	3
Medo = C	0	0	20	5	2	1	2
Neutro = D	0	0	2	25	0	0	3
Raiva = E	0	6	0	1	19	0	4
Surpresa = F	0	0	1	0	0	29	0
Triste = G	0	1	1	4	2	0	22

individualizada. Foi utilizado um NoteBook ASUS® Intel® Core i5, RAM 4 GB e HD 500 GB.

É importante destacar que os experimentos foram avaliados em dois momentos distintos. A primeira avaliação teve como foco a análise do desempenho referente à acurácia alcançada na classificação com imagens diferentes das utilizadas para o treinamento do modelo baseado em EC. Já a segunda, verificou a satisfação dos usuários referente as soluções flexíveis através do player de música.

A Tabela II, referente à primeira avaliação, exibe a Matriz de Confusão alcançada pelo Comitê de Classificação para a classificação das imagens dos usuários. Foram utilizadas 30 imagens de cada emoção por participante, totalizando 210 imagens. Vale salientar que tais imagens não foram utilizadas para o treinamento dos modelos, ou seja, as imagens eram desconhecidas, e alcançaram níveis de acurácia de 79,04%, próximos aos níveis de acurácia alcançado pelo trabalho proposto por Mano et al. (2015) [32], com acurácia de 80,53%.

Na segunda avaliação, foi possível verificar a satisfação do grupo por meio do Formulário de Observação do Participante (ver Figura 6). Os usuários foram submetidos ao teste por 1 hora e, a cada 15 minutos, lhes era perguntado a sua satisfação, de 0 a 10, com a utilização do sistema. Pôde-se, então, observar que a satisfação dos usuários que interagiram com o sistema (grupo 1) foi superior a dos usuários que não interagiram com o player de música (grupo 2).

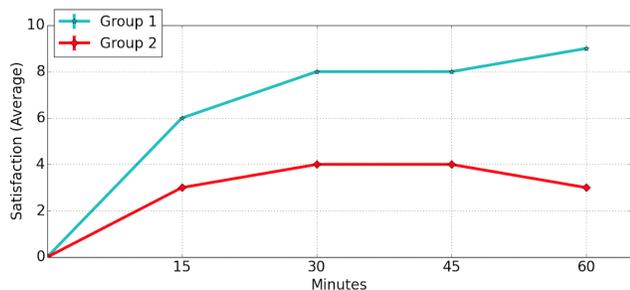


Fig. 6. Nível de satisfação de interação do usuário.

A partir dos experimentos realizados com os usuários, foi possível inferir que o uso de um sistema que se “adapte” as preferências dos usuários atende grande parte da intenção de identificar o estado emocional e sugerir músicas de acordo com a emoção do usuário.

VI. CONCLUSÃO

Na área de IHC inúmeros são os métodos, técnicas e instrumentos que apoiam a avaliação de respostas emocionais, devido à oportunidade de sistemas computacionais reagirem ao estado emocional de um indivíduo. Sendo assim, é necessário criar estratégias que suporte a captura, a análise e a classificação de indícios de respostas emocionais de usuários, permitindo que aplicações computacionais se flexibilizem a partir de tal classificação.

Diante disso, discute-se o uso de imagens de faces para a identificação e a classificação das emoções dos usuários de dispositivos computacionais. Este trabalho estabelece um sistema, baseado na classificação da face dos usuários, para interação com o usuário de acordo com suas emoções.

Os resultados mostraram que o Comitê apresentou baixa dispersão em sua replicação e a mediana da acurácia foi superior aos resultados apresentados pelos respectivos algoritmos que constituem o modelo proposto (usados individualmente) para a classificação da emoção do usuário quando comparado com abordagens clássicas. Ainda, pode-se observar também que a abordagem na utilização do player de música mostrou uma satisfação considerável dos usuários na interação afetiva de sistemas computacionais.

Vale ressaltar, portanto, que apesar dos bons resultados obtidos, acreditamos ser possível aumentar a satisfação na interação dos usuários. Assim, como trabalhos futuros, exploraremos: i.) outras abordagens para a identificação da emoção, por exemplo, o Componente de Tendências Comportamentais - Rede Social; ii.) os resultados da combinação de múltiplos sensores para a identificação da emoção e; iii.) estratégias de persuasão que melhor satisfazem as preferências dos usuários.

REFERENCES

- [1] F. Zhou, X. Qu, M. G. Helander, and J. R. Jiao, “Affect prediction from physiological measures via visual stimuli,” *International Journal of Human-Computer Studies*, vol. 69, no. 12, pp. 801–819, 2011.
- [2] V. P. Gonçalves, V. Neris, J. Ueyama, G. T. Giancristofaro, S. Seraphini et al., “Emoções na interação humano-computador: um estudo considerando sensores,” in *Proceedings of the 12th Brazilian Symposium on Human Factors in Computing Systems*. Brazilian Computer Society, 2013, pp. 252–255.
- [3] A. Lichtenstein, A. Oehme, S. Kupschick, and T. Jürgensohn, “Comparing two emotion models for deriving affective states from physiological data,” in *Affect and Emotion in Human-Computer Interaction*. Springer, 2008, pp. 35–50.
- [4] J. Klein, Y. Moon, and R. W. Picard, “This computer responds to user frustration:: Theory, design, and results,” *Interacting with computers*, vol. 14, no. 2, pp. 119–140, 2002.
- [5] J. N. Bailenson, E. D. Pontikakis, I. B. Mauss, J. J. Gross, M. E. Jabon, C. A. Hutcherson, C. Nass, and O. John, “Real-time classification of evoked emotions using facial feature tracking and physiological responses,” *International journal of human-computer studies*, vol. 66, no. 5, pp. 303–317, 2008.
- [6] C. Peter and B. Urban, “Emotion in human-computer interaction,” in *Expanding the Frontiers of Visual Analytics and Visualization*. Springer, 2012, pp. 239–262.
- [7] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, “Moodscope: Building a mood sensor from smartphone usage patterns,” in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013, pp. 389–402.

- [8] R. W. Picard, “What does it mean for a computer to “have” emotions? chapter to appear in: Emotions in humans and artifacts ed. by r. trappl, p. petta and s. payr,” 2004.
- [9] P. Eslambolchilar, M. Wilson, I. Oakley, and A. Dey, “Pinc: persuasion, influence, nudge & coercion through mobile devices,” in *CHI’11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2011, pp. 13–16.
- [10] H. Oinas-Kukkonen, “A foundation for the study of behavior change support systems,” *Personal and ubiquitous computing*, vol. 17, no. 6, pp. 1223–1235, 2013.
- [11] R. A. C. Xavier, F. E. Garcia, and V. P. de Almeida Neris, “Decisões de design de interfaces ruins eo impacto delas na interação: um estudo preliminar considerando o estado emocional de idosos,” in *Proceedings of the 11th Brazilian Symposium on Human Factors in Computing Systems*. Brazilian Computer Society, 2012, pp. 127–136.
- [12] A. Z. Saidin, C. Macaulay, and N. Hine, “Persuasion knowledge toolkit: requirements gathering with designer,” in *Proceedings of the 25th BCS Conference on Human-Computer Interaction*. British Computer Society, 2011, pp. 503–508.
- [13] K. R. Scherer, “What are emotions? and how can they be measured?” *Social science information*, vol. 44, no. 4, pp. 695–729, 2005.
- [14] S. Mahlke and M. Minge, “Consideration of multiple components of emotions in human-technology interaction,” in *Affect and emotion in human-computer interaction*. Springer, 2008, pp. 51–62.
- [15] G. Libralon and R. Romero, “Mapping of facial elements for emotion analysis,” in *Proceedings of the Brazilian Conf. on Intelligent Systems*, 2014.
- [16] J. R. Fontaine, Y. H. Poortinga, B. Setiadi, and S. S. Markam, “Cognitive structure of emotion terms in indonesia and the netherlands,” *Cognition & Emotion*, vol. 16, no. 1, pp. 61–86, 2002.
- [17] N. H. Frijda, *The emotions*. Cambridge University Press, 1986, no. 7.
- [18] B. Parkinson, *Ideas and realities of emotion*. Psychology Press, 1995.
- [19] K. R. Scherer, “Appraisal considered as a process of multilevel sequential checking,” *Appraisal processes in emotion: Theory, methods, research*, vol. 92, p. 120, 2001.
- [20] P. C. Ellsworth and K. R. Scherer, “Appraisal processes in emotion,” *Handbook of affective sciences*, vol. 572, p. V595, 2003.
- [21] P. Ekman and W. V. Friesen, “The repertoire of nonverbal behavior: Categories, origins, usage, and coding,” *Nonverbal communication, interaction, and gesture*, pp. 57–106, 1981.
- [22] G. Stemmler, “Methodological considerations in the psychophysiological study of emotion,” *Handbook of affective sciences*, pp. 225–255, 2003.
- [23] M. S. Yik, J. A. Russell, and L. F. Barrett, “Structure of self-reported current affect: Integration and beyond,” *Journal of personality and social psychology*, vol. 77, no. 3, p. 600, 1999.
- [24] P. Niedenthal, S. Krauth-Gruber, and F. Ric, “Psychology of emotion,” *Psychology of Emotion*, 2006.
- [25] J. A. Russell, “A circumplex model of affect,” *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.
- [26] P. Ekman, *Cross-cultural studies of facial expression*. New York: Academic, 1973.
- [27] A. Martinez and S. Du, “A model of the perception of facial expressions of emotion by humans: Research overview and perspectives,” *Journal of Machine Learning Research*, 2012.
- [28] P. Ekman, *Darwin and facial expression: A century of research in review*. Ishk, 2006.
- [29] L. Y. Mano, B. S. Façal, L. H. Nakamura, P. H. Gomes, G. L. Libralon, R. I. Meneguete, P. Geraldo Filho, G. T. Giancristofaro, G. Pessin, B. Krishnamachari *et al.*, “Exploiting iot technologies for enhancing health smart homes through patient identification and emotion recognition,” *Computer Communications*, 2016.
- [30] A. Campbell and T. Choudhury, “From smart to cognitive phones,” *Pervasive Computing, IEEE*, vol. 11, no. 3, 2012.
- [31] J. M. Saragih, S. Lucey, and J. F. Cohn, “Deformable model fitting by regularized landmark mean-shift,” *International Journal of Computer Vision*, vol. 91, no. 2, pp. 200–215, 2011.
- [32] L. Y. Mano, G. T. Giancristofaro, B. S. Façal, G. L. Libralon, G. Pessin, and P. H. Gomes, “Exploiting the use of ensemble classifiers to enhance the precision of user’s emotion classification,” in *Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS)*. ACM, 2015, p. 5.
- [33] P. Clark and T. Niblett, “The cn2 induction algorithm,” *Machine learning*, vol. 3, no. 4, pp. 261–283, 1989.
- [34] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.
- [35] G. Chanel, J. J. Kierkels, M. Soleymani, and T. Pun, “Short-term emotion assessment in a recall paradigm,” *International Journal of Human-Computer Studies*, vol. 67, no. 8, pp. 607–627, 2009.
- [36] A. Øhrn and T. Rowland, “Rough sets: a knowledge discovery technique for multifactorial medical outcomes,” *American journal of physical medicine & rehabilitation*, vol. 79, no. 1, pp. 100–108, 2000.
- [37] S. Ramakrishnan and I. M. El Emary, “Speech emotion recognition approaches in human computer interaction,” *Telecommunication Systems*, vol. 52, no. 3, pp. 1467–1478, 2013.
- [38] B. Schuller, S. Reiter, R. Muller, M. Al-Hames, M. Lang, and G. Rigoll, “Speaker independent speech emotion recognition by ensemble classification,” in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005, pp. 864–867.
- [39] A. J. Sharkey, “Multi-net systems,” in *Combining artificial neural nets*. Springer, 1999, pp. 1–30.
- [40] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.
- [41] J. Ueyama, V. P. Pinto, E. R. Madeira, P. Grace, T. M. Jonhson, and R. Y. Camargo, “Exploiting a generic approach for constructing mobile device applications,” in *Proceedings of the Fourth International ICST Conference on COMMUNICATION SYSTEM SOFTWARE AND MIDDLEWARE*. ACM, 2009, p. 12.