

Análise automática de reações emocionais aplicada a narrativas audiovisuais

Elton Cardoso do Nascimento¹, Artemis Moroni²

¹Instituto de Computação (IC)
Universidade Estadual de Campinas (Unicamp) - Campinas – SP – Brazil

²Divisão de Sistemas Ciberfísicos (DISCF)
CTI/MCTI Renato Archer - Campinas – SP – Brazil

elton.nascimento@students.ic.unicamp.br, artemis.moroni@cti.gov.br

Abstract. *The area of automatic emotion detection lately advances through detectors with greater accuracy and performance, and also in the actual use of these detectors in applications. Specifically, it is necessary to understand how the emotion estimated by these detectors is related to the stimuli, the environment in which the user is, and what relevant information it is possible to generate for the applications. Within this context, the objective of this work is to explore and make initial observations on how to use dimensional emotion detectors (valence-excitation) to automatically measure the reaction of people to events that happen in an audiovisual work, in this case an animated short film, collected using a camera. It is concluded that emotion detectors can generate significant information about the user's emotional reaction to the content. This type of analysis can advance and enable interactive applications that react to user emotion, such as installations, games, robots and human-computer interfaces in general.*

Resumo. *A área de detecção automática de emoções avança ultimamente através de detectores com maior acurácia e performance, e também no uso de fato desses detectores em aplicações. Em específico, é preciso entender como a emoção estimada por esses detectores se relaciona com os estímulos, o ambiente em que o usuário está, e quais informações relevantes é possível gerar para as aplicações. Dentro desse contexto, o objetivo desse trabalho é explorar e realizar observações iniciais sobre como utilizar detectores de emoção dimensional (valência-excitação) para mensurar automaticamente a reação de pessoas a eventos que acontecem em uma obra audiovisual, neste caso um curta animado, coletadas utilizando uma câmera. Conclui-se que detectores de emoção podem gerar informações significativas sobre a reação emocional do usuário ao conteúdo. Tal tipo de análise pode avançar e possibilitar aplicações interativas que reagem a emoção do usuário, como instalações, jogos, robôs e interfaces humano-computador em geral.*

1. Introdução

Dentre os modelos de inteligência artificial sendo desenvolvidos ultimamente existem

os detectores de emoção, que visam detectar e informar a emoção mostrada por uma pessoa, seja de forma categórica (ex. feliz, triste, neutro), ou de forma dimensional (ex. valência, excitação). Além de detectar a emoção, é preciso também pensar em formas de usar esses dados em aplicações, como pesquisas psicológicas ou entretenimento.

Quando se pensa nessas aplicações, torna-se interessante pensar não apenas na emoção do usuário, mas em quais estímulos que podem ter gerado tais reações. Muitos estudos focam apenas na parte de detecção, e a anotação e relação automática da reação com eventos ainda não foi muito explorada. A anotação automática da reação de eventos pode gerar informações importantes para aplicações como instalações artísticas e jogos eletrônicos, que podem, por exemplo, manipular sua narrativa segundo a reação do usuário para gerar a experiência estética almejada. Também pode ser utilizada pela interação humano computação e robótica, criando interfaces e agentes reativos ao usuário.

Neste trabalho, mostrarei explorações iniciais simples de como usar detectores de emoções para gerar informações sobre a reação emocional de uma pessoa enquanto assiste a uma obra audiovisual (um curta animado), utilizando também extratores automáticos de dados sobre filmes para anotar os eventos que acontecem durante a história. As principais contribuições são:

- Levantamento de fatores relevantes na análise de reação e uso dessa análise para entender como uma pessoa interage com uma obra audiovisual.
- Proposta de formas automáticas de registrar eventos e pontuar a reação do usuário.
- Proposta de um possível método para analisar reações longas/tônicas.

O artigo se divide da seguinte forma: a seção 2 mostra trabalhos relevantes para esse texto, dividido em 2.1, detecção de emoção, e 2.2, extratores de dados sobre filmes. A seção 3 discute possíveis questões éticas desse assunto. A seção 4 apresenta o método utilizado para explorar a área. A seção 5 apresenta e discute os resultados. Por fim, a seção 6 mostra as conclusões deste trabalho; enquanto a seção 7 contém as referências bibliográficas.

2. Trabalhos relacionados

Para esse estudo, foi relevante procurar por trabalhos já existentes de detecção e triangularização de emoções, que ajudaram a estruturar a metodologia e que serão descritos na primeira metade desta seção. Para gerar dados a partir do curta animado, precisei também procurar por modelos para extrair dados de filmes, mostrados no final desta seção.

2.1. Detecção e triangularização de emoção e seus usos

A primeira questão envolvida na detecção de emoções é como representá-las, sendo que a forma ideal pode variar segundo a aplicação. Podemos dividir os modelos entre modelos categóricos, que tentam separar as emoções em categorias discretas (alegria, tristeza, raiva, medo); dimensionais, que modela as emoções em um conjunto de dimensões, frequentemente valência (agradável – desagradável), ativação/excitação

(quão provável é da pessoa realizar uma ação dado o estado emocional), e controle (senso de controle sobre a emoção); e combinações dessas abordagens [Noroozi et al., 2018]. Existem diversas técnicas para detectar emoções, como utilizar dados fisiológicos, gestos corporais, ou expressões faciais.

Entre os modelos que utilizam expressões faciais, o “emonet” [Toisoul et al., 2021a] é um modelo dimensional que se destaca por ser recente, com modelo disponível publicamente¹, com métricas no estado da arte dessa área e que se propõe a detectar emoções não apenas no contexto de laboratórios, mas em contextos naturais. Como discutido por Toisoul et al (2021a) e Noroozi et al. (2018), é importante ressaltar que esses tipos de modelos podem ter vieses em relação às formas como emoções são expressadas por diferentes culturas e gêneros, além de captarem a emoção visível de uma pessoa, não seu estado interno. O emonet em específico é treinado utilizando dados de pessoas americanas-europeias, que podem não funcionar corretamente com pessoas latino-americanas, mais especificamente brasileiras.

Além da aferição de emoções, a interpretação dos dados também é uma tarefa importante. Nogueira et al. (2015) propõe uma ferramenta para automaticamente triangular emoções com eventos de jogos. Ele divide as reações emocionais baseando-se em definições da neurociência: reações fásicas, que possuem resposta rápida e curta, informando sobre a alteração de um estímulo; e reações tônicas, com resposta lenta e longa, informando sobre a presença e força de um estímulo [Binder et al. 2009]. A reação fásica é para um evento é definida como o máximo/mínimo global dos dados após filtragem, considerando uma janela ao redor de um evento, visto que a técnica utilizada para detectar emoções utiliza dados fisiológicos, que possuem latência maior que outras técnicas. Não é proposto no trabalho uma forma de detectar reações tônicas.

2.2. Análise de filmes

Utilizando técnicas de inteligência artificial, como redes convolucionais, é possível analisar e extrair dados de filmes automaticamente. É isso que realiza a MovieNet [Huang et al. 2020a], projeto que analisou de mais de mil filmes, treinando modelos para tarefas como detecção de atores, extração de *features* de atores (utilizado para identificar um mesmo ator entre detecções), detecção de ações e lugares, além de segmentação de cortes e cenas. Os modelos estão disponíveis publicamente no repositório do Github “movienet-tools” [Huang et al. 2020b].

3. Metodologia

Para investigar como os detectores de emoção podem ser utilizados para entender a reação de pessoas ao conteúdo sendo apresentado, realizei um experimento inicial para coletar primeiras observações. Foi escolhido um vídeo narrativo, um curta animado por possuir menor duração, personagens e eventos; e pessoas foram convidadas a assistir ao curta enquanto seus rostos eram gravados. O curta escolhido foi “A banda de um homem só” [Jimenez et al., 2005]. Por ser um trabalho inicial, foram coletados registros de apenas duas pessoas.

1 Modelo disponível no repositório do Github “Estimation of continuous valence and arousal levels from faces in naturalistic conditions, Nature Machine Intelligence 2021” [Toisoul et. all, 2021b]

Após coletar as gravações, elas foram processadas para obter dados de valência e excitação pelo modelo “emonet”, apresentado na seção 2.1 e escolhido pelas suas vantagens apresentadas, principalmente disponibilidade de código e por atuar em situações naturais, o que pode tornar possível no futuro utilizá-lo em instalações e outros contextos. Além dos dados das gravações, também foram gerados dados a partir do curta, utilizando os modelos da MovieNet, descrito anteriormente na seção 2.2. Mesmo sendo modelos treinados utilizando filmes com atores, tentei utilizá-los para extrair a posição e ação de personagens do curta.

Com as regiões que continham personagens extraídas, elas foram traqueadas utilizando o algoritmo “SORT” [Abewley, 2021], que busca entre os *frames* quais *bounding boxes* pertencem ao mesmo personagem. Do conjunto de detecções associado ao mesmo personagem é então utilizado um extrator de *features* para obter um vetor característico desse personagem, que é comparado através de um produto interno com o vetor extraído anteriormente de cada um dos personagens conhecidos do curta. Caso a comparação exceda um limiar, é associado a essa detecção o personagem; caso contrário a detecção é descartada, corrigindo erros de detecção do modelo. Os momentos em que um mesmo personagem aparece no curta foram reunidos em eventos, com o tempo inicial sendo o primeiro *frame* em que o personagem aparece, e o tempo final o último *frame* sequencial em que ele aparece. O conjunto de eventos é então filtrado segundo a duração, para reduzir a análise necessária e remover erros. Como os dados da ação que cada personagem realizava não se mostraram significativos, ou por serem muito genéricos, ou por estarem incorretos, eles não foram utilizados.

Os dados emocionais foram sincronizados manualmente com o curta e analisados qualitativamente de dois métodos diferentes: o primeiro manual, para gerar primeiras ideias sobre esses dados, em que se procurou pontos de destaque na reação emocional dos indivíduos, e tentou-se relacionar com momentos do curta e com a gravação facial. O segundo foi realizado com um processo automático, e depois mais uma vez com uma análise manual. O algoritmo primeiro filtra os dados utilizando um filtro de convolução, e após tenta associar quatro pontuações de reação para cada um dos eventos registrados, uma fásica e outra tônica, divididas em valência e excitação.

A reação fásica, semelhante ao trabalho de Nogueira et al. (2015), é definida como o maior pico em módulo da reação emocional (Equação 3). Diferente do trabalho citado, que considera um evento como um momento específico (sem duração), aqui os eventos são definidos a partir de seu momento de início e fim. Além disso, a técnica utilizada para obter os dados emocionais não possui a latência característica de modelos que utilizam fatores fisiológicos. Por esses fatores, foi analisado apenas os dados dentro do limite de tempo dos eventos, sem considerar dados antes ou após o evento. Assim como o trabalho de referência, apenas reações que se destacam da média das reações durante o evento são consideradas, removendo as reações dentro do intervalo $[\mu + \sigma, \mu - \sigma]$, como mostram as equações 1 e 2.

$$valência'(t) = \begin{cases} 0, & \text{caso } \mu_{valência} + \sigma_{valência} > valência(t) > \mu_{valência} - \sigma_{valência} \\ valência(t), & \text{caso contrário} \end{cases}$$

Equação 1: Filtragem da valência para cálculo da reação tônica

$$excitação'(t) = \begin{cases} 0, & \text{caso } \mu_{excitação} + \sigma_{excitação} > excitação(t) > \mu_{excitação} - \sigma_{excitação} \\ excitação(t), & \text{caso contrário} \end{cases}$$

Equação 2: Filtragem da excitação para cálculo da reação tônica

$$reação_{física}(evento) = (\max_{t \in evento} (|valência'(t)|), \max_{t \in evento} (|excitação'(t)|))$$

Equação 3: Definição da pontuação para reação física

O mensurador de reações tônicas proposto se baseia no fato de que essas reações são longas e estáveis, o que significa que a área delimitada pelos dados possivelmente é maior que em reações físicas, em que os dados variam rapidamente. Dessa forma, definiu-se a pontuação da reação como a integral da valência/excitação delimitada pelo evento (Equação 4), já que essa mesuraria a área sob o sinal, dividido pela duração do evento para manter a pontuação em um intervalo entre -1 e 1. Para desconsiderar dados constantes que não se relacionam com o evento, mais uma vez são desconsiderados dados próximos da média, porém nesse momento considerando a média e desvio padrão para a partir de alguns instantes antes e após o evento (2 s). Sequências inteiras com um desvio padrão muito pequeno (< 0,09) também são desconsideradas, com a reação sendo definida como 0.

$$reação_{tônica}(evento) = \left(\frac{\int_{evento} valência'(t) dt}{duração}, \frac{\int_{evento} excitação'(t) dt}{duração} \right)$$

Equação 4: Definição da pontuação para reação tônica

Por fim, para verificar a utilidade dessas pontuações, os quatro eventos com maior pontuação de cada participante (2 tônicas e 2 físicas divididas entre a de valência e excitação) foram manualmente analisadas em relação aos momentos do curta e a gravação dos participantes.

Todas as análises computacionais realizadas estão disponíveis no repositório do Github: <https://github.com/EltonCN/Emotion-estimation-applied-to-storytelling-reaction>. Ele não contém o curta utilizado e o modelo “emonet”, por questões legais; nem as gravações dos participantes, por questões éticas. Ainda assim, os códigos podem ser úteis para aplicações futuras.

4. Resultados e análise

Os eventos extraídos e relacionados aos personagens do curta podem ser vistos na Figura 1, enquanto a Figura 2 mostra esses eventos após a filtragem por duração, que removeu eventos com duração menor que 40 frames (1,6 s). É importante notar que, no curta, a primeira aparição do personagem 3 acontece após a primeira aparição do personagem 2, que por sua vez acontece após o personagem 1. Dessa forma, percebemos que, de fato, a filtragem conseguiu remover detecções incorretas, e os modelos da MovieNet se mostraram capazes de serem utilizados com animação. Porém, algumas aparições de fato dos personagens foram descartadas no processo.

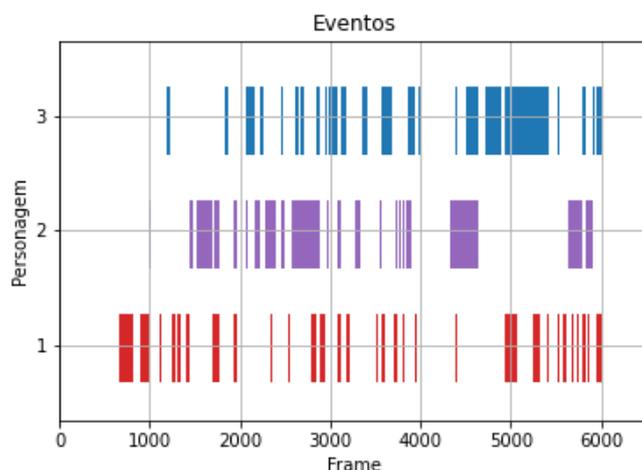


Figura 1: Eventos extraídos do curta

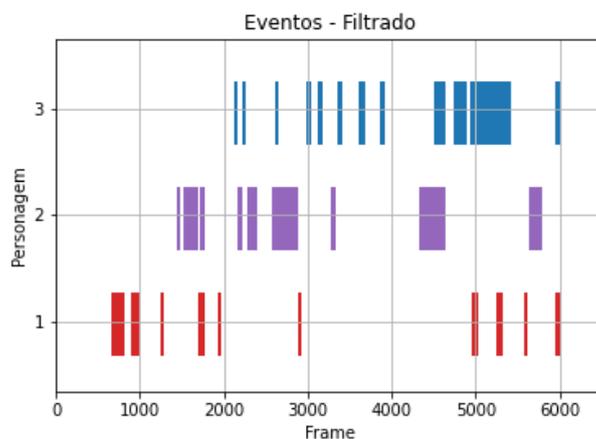


Figura 2: Eventos extraídos do curta após filtragem

Os dados de valência e excitação extraídos dos participantes se encontram nas figuras 3 e 4, respectivamente. Em um primeiro olhar, se destaca como as reações são diferentes para as duas pessoas (coeficiente de correlação 0,003 para a valência e 0,032 para a excitação entre os dois participantes), mesmo estando assistindo ao mesmo conteúdo. Um outro ponto importante é a reação logo após o frame 5000 do segundo sujeito, que se destaca em relação aos outros momentos dos dois participantes, se tratando de uma reação tônica, por ser estável e prolongada. A Figura 5 contém uma seção ampliada dos dados para mostrar o processo de filtragem.

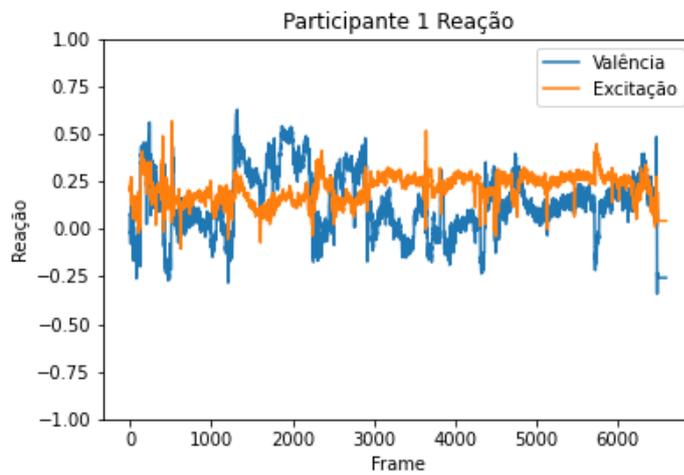


Figura 3: Reação do participante 1

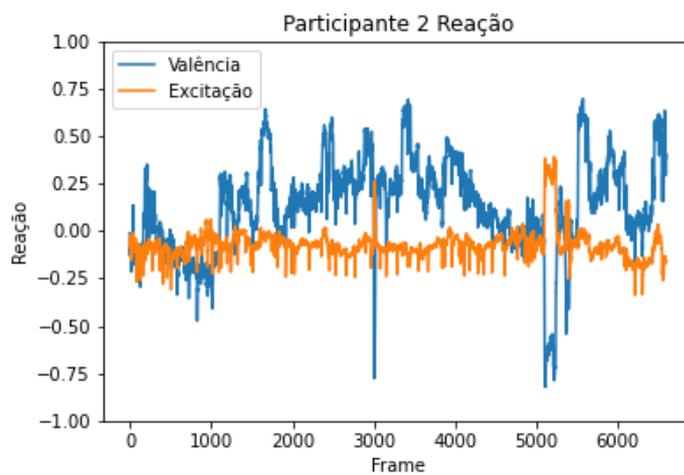


Figura 4: Reação do participante 2

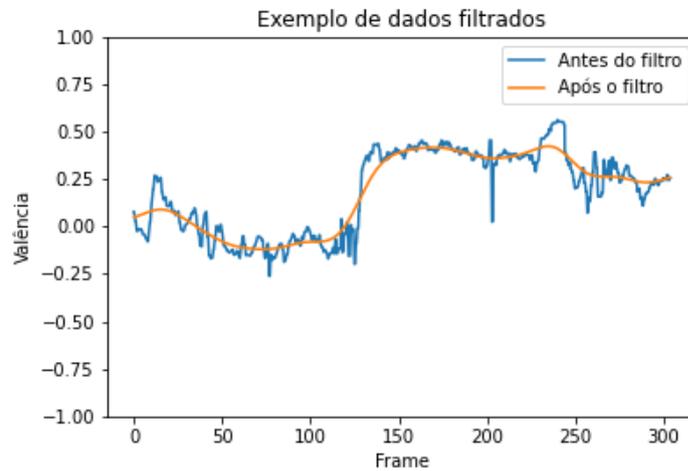


Figura 5: Exemplo de dados filtrados pela convolução

As figuras 6 e 7 mostram as reações dos sujeitos para cada evento, proporcionalmente dentro da mesma categoria tipo de reação – dimensão. Nota-se como as reações fásicas são mais comuns que as reações tônicas, sendo que uma das sequências (Participante 1 – Excitação) não apresenta nenhuma reação que passe pelo critério $\sigma > 0,09$. O fato da detecção depender da reação estar contida no evento sendo considerado pode impactar nisso, mas, ao observar as reações (figuras 3 e 4), é perceptível que em poucos momentos a reação se torna constante.

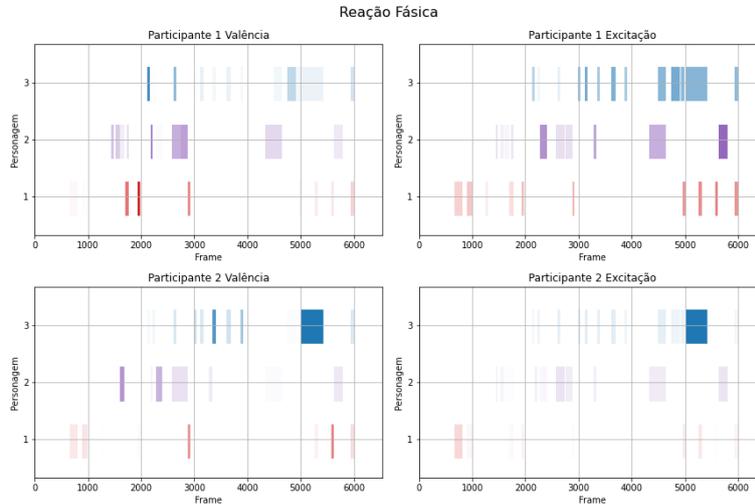


Figura 6: Eventos ponderados pela reação fásica, a opacidade indica a intensidade da reação em proporção com os outros eventos

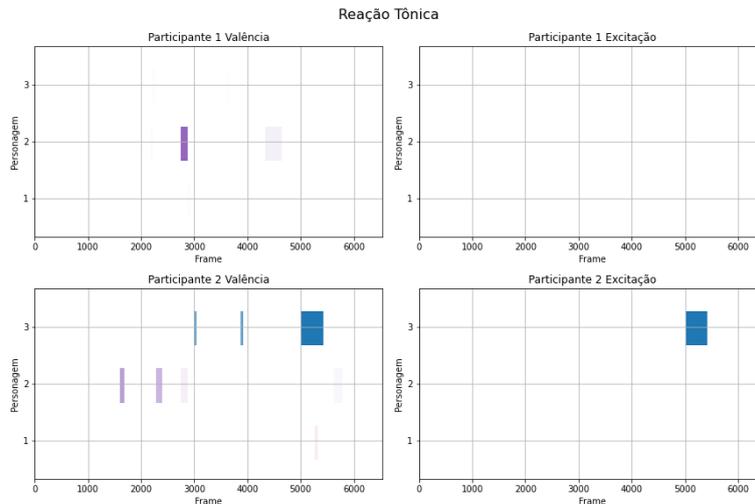


Figura 7: Eventos ponderados pela reação tônica, a opacidade indica a intensidade da reação em proporção com os outros eventos

Observando as reações em relação aos dados, a Figura 8 mostra em destaque os eventos com maior reação fásica, em que podemos perceber que as maiores reações estão ligadas aos maiores picos, como era esperado pelo método utilizado. Apenas na sequência participante 1 – valência isso não ocorre, pois o maior pico não está contido em nenhum evento.

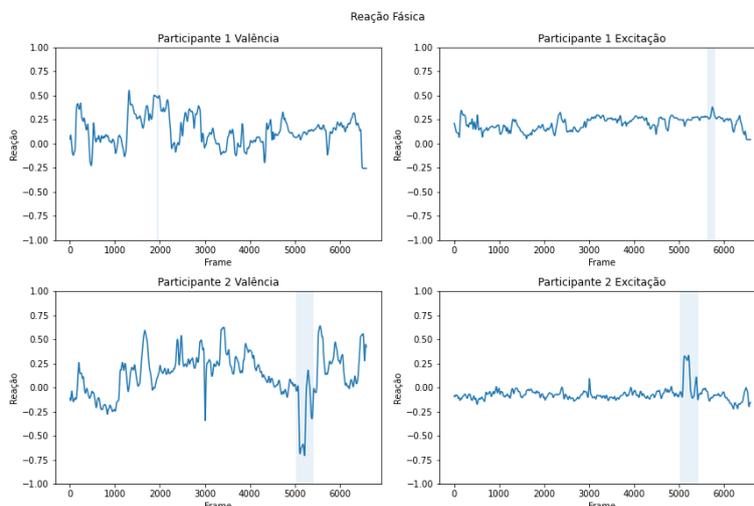


Figura 8: Reação física - Áreas destacadas indicam o evento de maior reação

Em relação à reação tônica, a Figura 9 também mostra as maiores reações em relação aos eventos. Nesse caso, é interessante observar como a reação tônica de valência do primeiro sujeito não é muito estável e prolongada como era desejado, sendo que o trecho no gráfico logo antes do frame 2000 possivelmente seria uma reação tônica mais característica, porém não é contida por nenhum evento. Mesmo assim, o trecho detectado é levemente estável, e se destaca das redondezas. Com o mesmo participante, mas em relação a excitação, não tivemos nenhuma reação, possivelmente porque, de fato, o participante não teve nenhuma reação tônica, visto que todos os momentos estáveis longos se aproximam da média. Por fim, as reações tônicas do segundo participante são as mesmas, tanto na valência quanto na excitação, e se tratam do evento que contém a reação tônica mais característica no gráfico, como já discutido anteriormente.

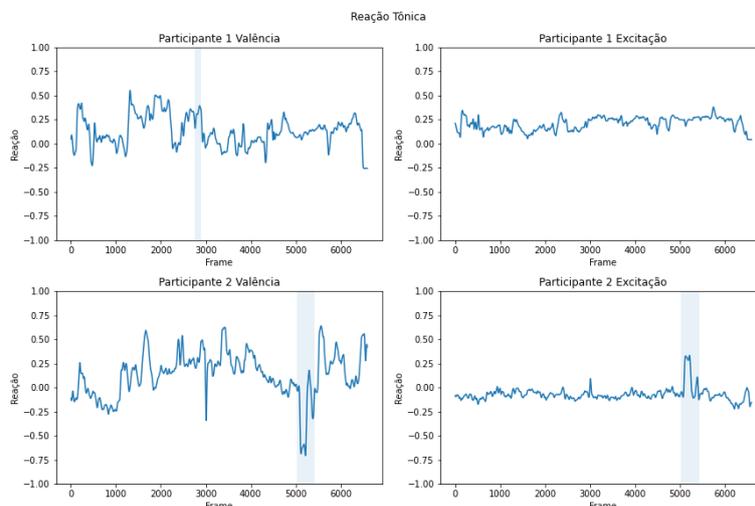


Figura 9: Reação tônica - Áreas destacadas indicam o evento de maior reação

Analisando e relacionando as reações com o curta e as gravações, primeiro a análise inteiramente manual a partir de trechos de interesse no gráfico, a Tabela 1 mostra alguns dos trechos analisados, mostrando também a emoção categórica determinada pela emonet. A primeira reação se trata de algo que acontece no início da primeira gravação, em que a justificativa aparenta ser o sujeito perceber que já assistiu ao curta, mostrando como é preciso considerar que nem todas as reações são diretamente estimuladas por eventos que acontecem em tela. A segunda e terceira observação se tratam do mesmo momento no curta, porém com reações diferentes para cada sujeito. Neste momento, uma das personagens oferece sarcasticamente moedas para os músicos interesseiros que acabaram de ocasionar em seu sofrimento, para depois jogá-las em uma fonte d'água. Observando os dados, é possível pensar que, enquanto o primeiro sujeito aparenta ter se focado em julgar a ação dos músicos, o segundo se foca no sarcasmo visível da personagem, prevendo o que aconteceria em seguida. O penúltimo trecho se trata da reação em destaque do sujeito dois já mencionada anteriormente, durante ele a garota começa a tocar um violino, tocando muito desafinadamente, o que causa repulsa ao participante, e que termina assim que a garota começa a tocar corretamente o instrumento, no último trecho analisado. Isso indica que, não apenas os eventos visuais, mas também os sonoros são relevantes. Essa primeira análise mostra quão representativos são os dados gerados por esses modelos em relação à reação dos participantes. Mesmo não tendo sido treinado com muitos dados de pessoas brasileiras, os dados não aparentam ter dificuldade em representar as reações.

Tabela 1: Reações interessantes observadas manualmente

Participante	Valência	Excitação	Emoção categórica	Informações relevantes do curta ou gravação
1	0,63	0,25	Feliz	Pessoa percebe já ter assistido ao curta anteriormente

1	-0,08	0,45	Raiva	Garota oferece moedas aos músicos
2	0,17	-0,11	Feliz	Garota oferece moedas aos músicos
2	-0,39	0,13	Desgosto	Garota toca o violino desafinado
2	0,06	-0,06	Neutro	Garota começa a tocar bem o violino

Partindo para a análise manual relacionando as reações com maiores pontuações com o curta e as gravações, a Figura 10 mostra a representação circunflexa dessas reações detectadas, enquanto a Tabela 2 sumariza as relações encontradas. Como as pontuações geradas, principalmente para as reações tônicas, são muito pequenas, a Figura 10 contém essas representações com as pontuações normalizadas para caírem sobre a borda do círculo, melhorando a visualização. Mais uma vez foi possível relacionar as reações com os momentos do curta. A primeira reação se refere a um trecho cômico do curta, e de fato a reação do participante indica isso. A segunda reação (participante 1 - fásica - valência) é interessante. Enquanto a análise anterior mostrou esse evento gerando uma reação de raiva, pela interpretação circunflexa a pontuação gerada para esse evento seria de algo entre “animação” e “atenção”, com sinais contrários na dimensão de valência. Observando melhor a região em que esse evento acontece (Figura 11), é possível entender porque isso acontece: enquanto os dados brutos indicariam nesse evento que a maior reação (maior pico) seria negativo, ao suavizar os dados o maior pico se torna positivo. Isso pode ser de fato um problema do filtro, porém poderia ter sido corrigido se a maior pontuação tivesse sido comparada junto com o valor na outra dimensão no mesmo instante do pico, visto que o pico em excitação ocorre no mesmo momento que o pico de valência esperado para esse evento.

O próximo evento (participante 1 – tônica – valência) ocorre em continuação a um evento cômico anterior, o que justifica a pontuação de valência tônica, porém é interessante observar como a reação do sujeito se altera durante o evento. Por fim, o único evento analisado para o sujeito 2 corresponde ao evento de maior destaque mencionado anteriormente e analisado manualmente, com as pontuações tônica e fásica alinhadas (Figura 10). A representação circunflexa (chateado-estressado) se alinha com a gerada pela emonet (desgosto). Um detalhe que não é sustentado de nenhuma forma pela análise automática é o fato da garota tocar o violino desafinadamente dois instantes seguidos, e a reação facial do sujeito responde a esses dois instantes. Observando novamente os dois gráficos correspondentes da Figura 9, é possível perceber que a reação tônica se inicia e termina com picos, que correspondem a esses instantes.

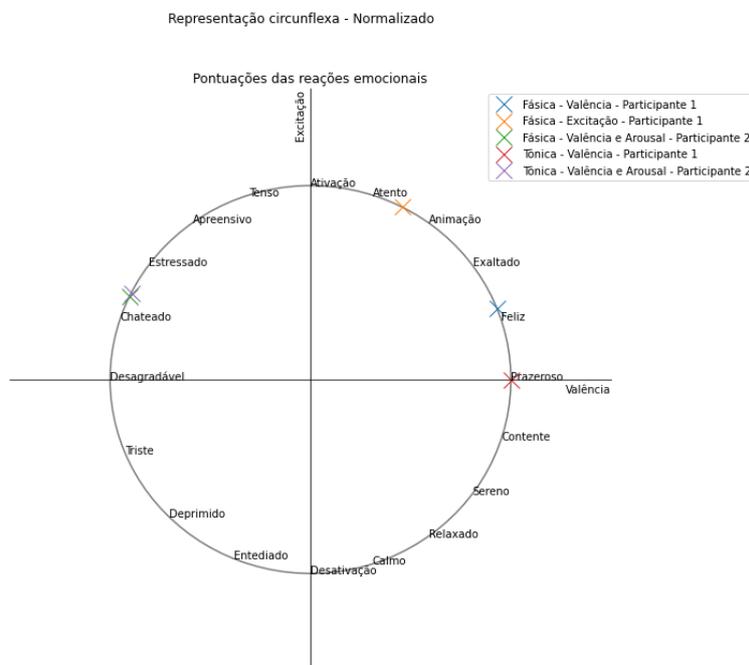


Figura 10: Representação circunflexa normalizada das reações com maiores pontuações

Tabela 2: Análise das reações com maior pontuação

Participante	Tipo de reação	Pontuação Valência	Pontuação Excitação	Momento do curta	Momento da gravação
1	Física Valência	-0,50	0,20	Garota ouve pela primeira vez o músico 2, e começa a andar em direção ao seu estande	Início de uma risada do fim do evento
1	Física Excitação	-0,19	0,38	Garota oferece moedas aos músicos	O sujeito está focado assistindo.
1	Tônica Valência	-0,09	0,00	Logo antes do evento, o músico 1 cai com o instrumento (momento cômico). Durante o evento, o músico 2 toca o violino, porém começa a ser interrompido pelo músico 1 com uma bateria	O sujeito inicia o evento sorrindo, mas altera para uma face mais séria durante o evento
2	Física Valência e excitação	-0,71	0,33	Garota toca o violina desafinadamente duas vezes antes de começar a tocar	Face apresenta duas vezes desgosto durante a primeira metade do evento

	Tônica Valência e excitação	-0,19	0,09	corretamente	
--	-----------------------------------	-------	------	--------------	--

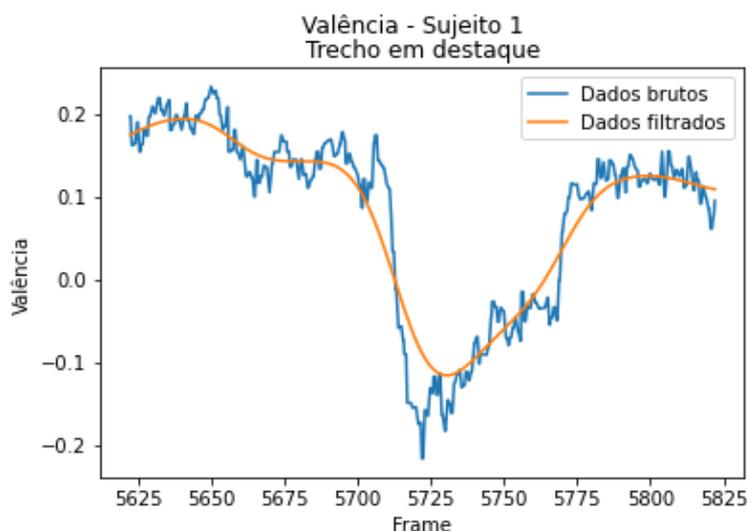


Figura 11: Região com contradição nas análises em destaque

Com todas essas análises, podemos pensar em como esses dados poderiam ser utilizados, por exemplo, em um sistema de uma instalação interativa ou jogo eletrônico. Considerando o momento em que a garota toca desafinadamente o violino, isso poderia ser utilizado por alguma sequência que tem como objetivo causar repulsa ao jogador para o segundo participante, enquanto para o primeiro seria preciso procurar algum outro estímulo. O momento em que a garota oferece as moedas aos músicos poderia ser utilizado por um jogo narrativo para criar ramificações na história, uma mais agressiva contra os músicos para o primeiro jogador, e outra não tanto.

Por fim, é importante ressaltar que as análises realizadas computacionalmente foram todas offline, após realizar as gravações e sem uma restrição de tempo. Portanto, não houve muitos esforços para otimizar ou avaliar a performance dos algoritmos. Para uma aplicação em tempo real, é preciso considerar o tempo necessários para executar as redes neurais, sendo que a “emonet” executando em uma placa gráfica “RTX 2060” conseguiu uma performance por volta de 60 fps, desconsiderando o tempo necessário para movimentar os frames da CPU até a GPU. Os modelos da MovieNet, que podem não ser necessários dependendo da aplicação, superaram uma hora de execução para o curta de aproximadamente 4 min, mas é preciso considerar que o decodificador de vídeo da biblioteca OpenCV utilizado não é eficiente. Dentre os processos de pós-processamento, o mais custoso é a filtragem por convolução, que pode futuramente ser acelerado utilizando a GPU.

5. Conclusão

Interfaces reativas às emoções dos usuários ainda precisam de muitos estudos para se tornarem reais, desde como detectar e representar essas emoções através de dados do usuário; até como interpretar essas detecções e gerar programas que reagem a elas. Nesse artigo, mostrei um trabalho inicial que explorou como os dados de emoção podem ser interpretados para utilização futura, contribuindo com métodos para detectar e pontuar eventos, principalmente de forma tônica, e levantando os seguintes pontos que podem ser relevantes para aplicações reativas, que ainda precisam ser melhores explorados e comprovados em trabalhos futuros:

- Dados de detecção de emoção podem, de fato, gerar informações importantes sobre a reação emocional do usuário, principalmente se vinculados com eventos específicos.
 - Observar as reações entre usuários também pode ser útil.
 - Essas informações possivelmente podem ser utilizadas para criar instalações e jogos reativos à emoção, como narrativas ramificadas.
- No contexto de obras audiovisuais, tanto os estímulos visuais quanto sonoros são importantes para entender a reação do usuário.
- O contexto e histórico do usuário pode influenciar sua reação, como ele ter familiaridade com o conteúdo.
- É possível detectar e pontuar reações tônicas, que também possuem informações importantes sobre a reação do usuário.
 - As reações tônicas aparentam ser menos comuns que as fásicas.

Para trabalhos futuros, é possível continuar explorando não apenas esses pontos, mas também alguns problemas apresentados, como a performance para execução em tempo real, computar a reação fásica em valência e excitação de um mesmo instante, suportar múltiplas reações para um mesmo evento (é possível encontrar um exemplo disso em [Nogueira et al. 2015]). Considerar o inverso, procurar primeiro por reações e depois associar aos eventos, também pode ser utilizado para múltiplas reações por eventos, ou até mesmo para reações sem eventos associados. Possivelmente utilizar eventos mais representativos, como eventos gerados automaticamente pelas aplicações para os diferentes estímulos, também pode ajudar a melhorar a análise. É preciso ainda também pensar em como criar sistemas que utilizam essas informações de forma ética, considerando quais devem ser as restrições desses sistemas. Explorar como desenvolver aplicações reativas em interface humano-computador, robótica e arte tecnológica também pode ser um passo futuro.

6. Referências

- Noroozi, F., Corneanu, C. A., Kamińska, D., et al. (23 jan 2018). Survey on Emotional Body Gesture Recognition. arXiv:1801.07481 [cs],
- Toisoul, A., Kossaifi, J., Bulat, A., Tzimiropoulos, G. and Pantic, M. (jan 2021a). Estimation of continuous valence and arousal levels from faces in naturalistic conditions. *Nature Machine Intelligence*, v. 3, n. 1, p. 42–50.

- Toisoul, A. and Kossaifi, J. (2021b). Estimation of continuous valence and arousal levels from faces in naturalistic conditions, *Nature Machine Intelligence* 2021. face-analysis. Disponível em: <https://github.com/face-analysis/emonet>. Acesso em: 22 ago, 2021.
- Nogueira, P. A., Torres, V., Rodrigues, R. and Oliveira, E. (1 jun 2015). An annotation tool for automatically triangulating individuals' psychophysiological emotional reactions to digital media stimuli. *Entertainment Computing*, v. 9–10, p. 19–27.
- Binder, M. D., Hirokawa, N. and Windhorst, U. [Eds.] (2009). Adaptation of Sensory Receptors. *Encyclopedia of Neuroscience*. Berlin, Heidelberg: Springer. p. 54–54.
- Huang, Q., Xiong, Y., Rao, A., Wang, J. and Lin, D. (21 jul 2020a). MovieNet: A Holistic Dataset for Movie Understanding.
- Huang, Q., Xiong, Y., Rao, A., Wang, J. and Lin, D (2020b). movienet-tools. Disponível em: <https://github.com/movienet/movienet-tools>. Acesso em: 22 ago, 2020.
- A BANDA de Um Homem Só. Direção de Andrew Jimenez, Mark Andrews. Produção de Osnat Shurer, John Lasseter, Brad Bird. Roteiro: Andrew Jimenez, Mark Andrews. 2005. (4 min.), son., color.
- Abewley (2021). SORT. Disponível em: <https://github.com/abewley/sort>. Acesso em 22 ago, 2021.