

**ANÁLISE SEMIODISCURSIVA DE NÍVEIS DE TENSÃO EM VÍDEOS  
TELEVISIVOS POR MEIO DA INFERÊNCIA AUTOMÁTICA DE EMOÇÕES**  
SEMIODISCURSIVE ANALYSIS OF TENSION LEVELS IN TV VIDEOS BASED  
ON THE AUTOMATIC INFERENCE OF EMOTIONS

Moisés Henrique Ramos PEREIRA  
Instituto de Engenharia e Tecnologia (UNI-BH)  
[moises.ramos@prof.unibh.br](mailto:moises.ramos@prof.unibh.br)

Flávio Luís Cardeal PÁDUA  
Departamento de Computação (CEFET-MG)  
[cardeal@decom.cefetmg.br](mailto:cardeal@decom.cefetmg.br)

Giani DAVID-SILVA  
Departamento de Linguagem e Tecnologia (CEFET-MG)  
[gianids@deii.cefetmg.br](mailto:gianids@deii.cefetmg.br)

Resumo: Este artigo aborda o problema de análise semiodiscursiva de níveis de tensão em vídeos televisivos, por meio da inferência automática de emoções em expressões faciais de seus participantes (âncoras e repórteres de telejornais, por exemplo). A motivação maior para esse estudo consiste em contribuir com estudos inovadores sobre a determinação automática dos níveis de tensão de programas televisivos no Brasil e auxiliar na identificação de estratégias comunicativas desses programas. Nesse contexto, introduz-se a problematização que envolve conceitos de Linguística, em especial aqueles relacionados à semiótica tensiva de programas televisivos, e fundamentos da Ciência da Computação referentes ao uso de técnicas para reconhecimento de emoções, a partir dos conteúdos dos vídeos associados aos programas. A fim de fundamentar o trabalho desenvolvido, este artigo apresenta alguns dos principais trabalhos do estado da arte das áreas envolvidas, expondo-se em seguida um estudo de caso e a definição do *corpus* considerado neste trabalho. Uma visão geral da metodologia empregada é apresentada no que tange à conceituação dos elementos discursivos que constituem os níveis de tensão de vídeos televisivos e das técnicas computacionais para a detecção de faces e reconhecimento de expressões faciais. Finalmente, são apresentadas algumas regras de dedução para a classificação dos vídeos televisivos quanto a sua tensão discursiva que foram modeladas junto a analistas do discurso do Centro de Apoio a Pesquisas sobre Televisão (CAPTE) do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG).

**Palavras-chaves:** Níveis de Tensão; Inferência de Emoções; Expressões Faciais.

Abstract: This paper addresses the automatic determination of tension levels in TV programs by using image analysis techniques for emotion inference from facial expressions of video participants (such as, anchors and reporters in newscasts). The greatest motivation for this study is to contribute with innovative studies on the automatic determination of the television program tension levels in Brazil and help identifying communication strategies of these programs. In this context, the problem statement is introduced by considering concepts of Linguistics, in particular those related to tensile semiotics of television programs, and computer science fundamentals concerning the use of techniques for emotion recognition from the contents of the videos associated with the programs. A comprehensive literature review is presented, followed by the description of a case study and the definition of the *corpus* considered in this work. An overview of the methodology is presented regarding the concept of discursive elements that constitute the tension levels of television videos and computational techniques for face detection

and recognition of facial expressions. Finally, we present some deduction rules for the classification of television videos according to their discursive tension levels modeled by discourse analysts of the Center for Research Support on Television (CAPTE) of the Federal Center for Technological Education of Minas Gerais (CEFET-MG).

**Key-words:** Tension Levels; Emotion Inference; Facial Expressions.

## **Introdução**

Este artigo propõe uma nova abordagem para a análise semiodiscursiva de níveis de tensão em vídeos televisivos, por meio da inferência automática de emoções em expressões faciais de seus participantes. Quando se assiste a um programa televisivo, presencia-se, com certa frequência, uma manobra em preparar o modelo mental dos telespectadores para assimilarem melhor as informações que serão transmitidas, mesmo que isso não seja percebido pelo público (PIMENTEL, 2009). Essa premissa é evidente nos telejornais que exibem a chamada das manchetes e, logo a seguir, uma reportagem com conteúdo emocional moderado, variando a tensão emocional significativamente, à medida que mais reportagens são apresentadas durante o programa.

Percebe-se, também, que o sequenciamento das notícias a fim de transmitir credibilidade é conduzido, sob um estilo próprio, pelos telejornalistas (GOFFMAN, 1981). Além das técnicas de redação para afirmar o caráter objetivo do telejornal, acrescenta-se a contenção e o equilíbrio de movimentos gestuais e expressivos, por parte dos apresentadores, elementos que permitem ao programa tomar uma posição sobre aquilo que é noticiado e estabelecer uma identificação com o telespectador (DAVID-SILVA, 2005).

O estilo dos programas televisivos é construído, geralmente, pela redundância e, como a fala é única, o sujeito de informação escolhe, consciente ou inconscientemente, os recursos que irá utilizar em uma situação específica de comunicação, tais como modulações vocais, expressividade facial e gestos corporais (MADUREIRA, 2004).

Desde o início do telejornalismo mundial, com primeira notícia transmitida pela emissora WGY em agosto de 1928 nos Estados Unidos (CLARK, 1962), até a década de 1960, prevaleceu o modelo apresentador-ventríloquo, responsável apenas por ser o tão somente locutor das informações, ou seja, ele era meramente um porta-voz das informações. Normalmente, esse tipo de apresentador atinha-se ao texto (eram poucas as atitudes diferenciadas através de olhares, entonações de voz, gestos). Com o passar do tempo, o apresentador passou a ter função central nos telejornais, por meio dele é

que o telejornal se estrutura, estabelecendo-se o eixo Y-Y (olhos nos olhos). A enunciação é evidenciada, o apresentador fala para nós e nos olha ao falar. Assim, Véron (1983: 112-113) caracteriza esses dois tipos de apresentadores:

Apresentador ventríloquo – o corpo do apresentador está lá, mas a dimensão do contato é reduzida ao olhar. A gestualidade é anulada, postura rígida do corpo, feição em “grau zero”. A palavra é desprovida de todo operador de modalização, o texto dito (ou lido) é estritamente descritivo, factual. O espaço em torno do apresentado é também mínimo. Apresentador é um suporte neutro, um ponto de passagem (a informação fala pela sua boca).

Apresentador moderno – o encontro dos olhares se torna o eixo que sustenta a construção do “corpo” mediatizado do enunciador. Os operadores desse sistema modalizam a fala e constroem um elo com o telespectador. Os gestos podem criar uma distância em relação ao que é dito (reforçada pela distância lingüisticamente imposta) permite ao mesmo tempo de estabelecer uma identificação com o telespectador; “a distância calculada engendra a confiança, a crença” (VÉRON, 1983: 112-113).

O eixo Y-Y, então, associado a um movimento de referência, assemelha-se a uma operação destinada à desficcionalização do discurso. Se considerarmos que a realidade supõe a presença (espaço/tempo) e que só é plenamente real o aqui e agora (ODIN, 2000), o apresentador vai exercer a função de ancoragem do discurso no real da atualidade – um operador de realização (VÉRON, 1983), na tentativa de “neutralizar” ao máximo o estatuto ficcional inerente a todo discurso.

Nos últimos anos, a sequência das notícias e a articulação dos apresentadores de programas televisivos vêm sendo objetos de estudo. Existem diversos esforços em descobrir padrões que revelem a estratégia comunicativa pretendida por meio do sequenciamento de notícias, os diferentes planos fílmicos empregados e a influência dos apresentadores na construção do *ethos* discursivo do programa (SABINO, DAVID-SILVA & PÁDUA, 2014; PEREIRA *et al.* 2014). De forma natural, os jornalistas devem causar impacto, interpretar e despertar sentimentos diversos no telespectador por meio da expressividade corporal, sendo esses objetivos pautados pelas emissoras de televisão (GODOY-COTES, 2008; CHARAUDEAU, 2006).

Observa-se, então, que existe uma demanda em realizar a análise discursiva de programas de televisão por meio de recursos audiovisuais que são identificados, *a priori*, em um processo manual de cruzamento de informações. Sobre esses recursos e os respectivos dados extraídos, ocorre o trabalho dos analistas do discurso em identificar a estratégia comunicativa, subentendida naqueles objetos de estudo. Neste

contexto, este artigo debruça-se na implementação de um processo automático da extração desses recursos, especificamente as expressões faciais, para a inferência de emoções em vídeos, a fim de subsidiar a análise dos níveis de tensão em programas televisivos.

### **Problematização**

No estudo sobre a mídia televisiva, percebe-se que o problema da tensão está na percepção semântica do telespectador sobre o fato e, dessa forma, trata-se de um problema subjetivo a ser modelado. Uma imagem em *close*, por exemplo, sugere uma tensão maior para uma determinada expressão facial, porque pretende-se envolver mais o espectador, mas não se pode afirmar se isso realmente ocorreu. Dessa forma, os esforços da Análise do Discurso concentram-se sobre as instâncias de produção, na busca por padrões que caracterizem os gêneros televisivos, assim como a identificação de efeitos que visam captar e fidelizar o seu interlocutor (CHARAUDEAU, 2006).

Dentre os padrões de estudos encontrados na literatura, analisam-se as expressões faciais dos apresentadores, as modulações da fala, os movimentos gestuais, a intensidade visual nos espaços do telejornal e os enquadramentos da câmera. Desses padrões, este artigo propõe-se a automatizar os processos de identificação e extração de expressões faciais, inferindo as respectivas emoções associadas por meio de gráficos que servirão como fontes de dados auxiliares para o trabalho do analista do discurso.

O reconhecimento de expressões faciais fornece, diretamente, suporte ao reconhecimento de emoções. Este artigo expõe alguns dos trabalhos mais relevantes na área de reconhecimento de expressões faciais e emoções com o intuito de empregar os melhores modelos e alcançar resultados compatíveis com o estado da arte, levando-se em consideração a movimentação dos indivíduos nos vídeos televisivos, incluindo telejornais. Os modelos devem ser robustos em relação aos espaços visuais que incluem ambientes mais controlados, tais como o estúdio do programa, e ambientes externos que, geralmente, formam o espaço das reportagens. Uma etapa crucial para a tarefa de reconhecimento de expressões faciais é a relacionada à detecção de faces dos indivíduos nos vídeos, pois as cenas são dinâmicas e os indivíduos se movem o tempo todo, mesmo que os movimentos, algumas vezes, sejam inexpressivos.

Para um processo de reconhecimento de expressões faciais coeso, devem-se mapear devidamente as respectivas emoções associadas e o nível de granularidade a ser

empregado. Foram encontrados diversos trabalhos que propunham grupos de expressões faciais em torno de seis emoções básicas recorrentes na literatura, quais sejam: raiva, medo, aversão, surpresa, alegria e tristeza (BETTADAPURA, 2009; EKMAN & FRIESEN, 1978). Definir como apresentar os níveis de tensão e as respectivas transições emocionais é um desafio instigante, enfrentado neste trabalho.

### **Estado da Arte**

No âmbito da Análise do Discurso, existem alguns trabalhos na literatura que se debruçaram em estudar a tensão e a sequência das temáticas apresentadas em programas televisivos, incluindo notícias de telejornais.

Na pesquisa realizada por Friedman, DiMatteo & Mertz (1980), as expressões faciais dos apresentadores de telejornais americanos foram analisadas durante a cobertura da campanha eleitoral presidencial de 1976. A possibilidade de comunicação não verbal nas notícias, mesmo que sutil e sistemática, foi explorada por meio da análise do conteúdo não verbal, no qual foram estudadas as expressões faciais que acompanharam a enunciação dos nomes dos candidatos. Os resultados mostraram implicações importantes no papel da comunicação não verbal, na transmissão de informações para os telespectadores, principalmente em relação às notícias com temáticas que possuíam conteúdo mais afetivo.

No estudo sobre a ordenação das notícias em telejornais quanto às temáticas, sob uma nova abordagem na área, David-Silva (2005) analisou quatro telejornais, dois brasileiros e dois franceses, encontrando semelhanças marcantes entre eles, no que se refere às temáticas abordadas e à sintaxe pela qual as notícias foram organizadas, tendo-se a tendência de ir de um ponto máximo de tensão, normalmente centrado em matérias que demonstram a desordem do mundo, para certa sensação de retomada da ordem ou leveza ao abordar notícias de conteúdo esportivo, de lazer, dentre outros assuntos. Com vasto levantamento das notícias, em diversas datas e extensos tempos de exibição, a autora determinou os níveis de tensão com base no conteúdo temático e no possível impacto emocional no telespectador, considerando-se os sentimentos envolvidos. Essa metodologia foi reaplicada nos estudos de Braighi-Andrade (2013) e de Silva (2014).

Os trabalhos realizados por Fechine (2008) e Butturi & Lenzi (2009) enfatizaram na revisão e na elaboração de estudos que tomem a tensão dos programas televisivos como fortes indicadores dos *ethos* dos espaços de enunciação, ou seja, qual “a imagem

de si” ou postura que o apresentador faz ao proferir o seu discurso, a fim de persuadir outrem acerca da verdade do que é enunciado.

Considerando os demais veículos de informação, além de vídeos, tem-se em Macgilchrist (2011) uma análise sobre a tensão no discurso das notícias sobre a Rússia após a dissolução da União Soviética. Este trabalho mostra que o jornalismo desempenha um papel mais incisivo na política e desenvolve uma abordagem para investigar discursos hegemônicos, fissuras discursivas e tensões, confirmando que a forma como as notícias são enunciadas evidencia que no jornalismo a instabilidade da ordem social torna-se visível para o público, por conta da tensão discursiva empregada.

No âmbito do reconhecimento de emoções em vídeos, considerando que a expressão de emoções ocorre principalmente pela face humana, o rosto pode ser considerado indispensável para comunicar diferentes mensagens e percepções, inclusive as emoções do locutor (EKMAN & FRIESEN, 1978). Os projetos a seguir contribuíram de forma significativa para as pesquisas na área e se basearam no uso de técnicas e descritores robustos para a detecção de faces em imagens e vídeos, bem como algoritmos para o reconhecimento de expressões faciais na inferência de emoções.

Em Ekman & Friesen (1978), os autores demonstraram evidências de que as expressões faciais de emoções podem ser inferidas por sinais rápidos da face. Estes sinais são caracterizados por mudanças na aparência da face que duram segundos ou frações de segundo, uns mais visíveis que outros. Com isso, os autores formularam o modelo de emoções básicas, fundamentado sobre seis expressões faciais (raiva, medo, repulsa, surpresa, alegria e tristeza) que são encontradas em diversas culturas e são exibidas da mesma forma, desde crianças até idosos. Foram mapeados os pontos de singularidade de cada tipo de expressão facial com base em testes realizados sobre um vasto banco de imagens, gerando um importante modelo usado por diversos trabalhos.

No trabalho de Viola & Jones (2001), os autores propuseram um detector de face rápido que analisa cada quadro do vídeo de forma independente dos anteriores, ou seja, considerando que cada quadro do vídeo fosse uma imagem isolada, sem lidar com a dinâmica da cena. Esse método detecta faces frontais nas imagens em nível de cinza, e as respectivas características são usadas para treinar um conjunto de classificadores.

O trabalho de Diniz (2013) propõe melhorias no processo de aprendizagem em um Ambiente Virtual de Aprendizagem (AVA), por meio de agentes inteligentes e

descreve um sistema em tempo real que utiliza técnicas de Análise dos Componentes Principais para reconhecimento facial; um Sistema de Codificação de Ação Facial para rastreamento dos pontos característicos faciais; e um método de classificação da emoção baseado em regras para reconhecimento de expressões faciais. Com auxílio de uma câmera, esse sistema permite identificar o aluno durante a realização das avaliações pedagógicas no AVA, além de reconhecer a emoção do aluno pela sua expressão facial, a fim de motivá-lo à aprendizagem de forma eficaz, utilizando um agente pedagógico animado.

Em cada expressão facial, diferentes partes do rosto desempenham papéis específicos, devido ao movimento distinto dos músculos faciais. No trabalho de Liao *et al.* (2013), foi estudado o aprendizado dos pesos associados a cada região facial para diferentes expressões. Os pontos de característica facial são inicialmente localizados com base em um modelo gráfico de precisão. Utilizando-se o fluxo óptico para representar a informação de movimento devido à expressão facial, um problema é formulado para o aprendizado, a partir de dados de treinamento em que faces de mesma categoria de expressão estivessem mais próximas. Foram demonstradas as vantagens da aplicação do método de aprendizado desses pesos para reconhecimento de expressões faciais e as respectivas estimativas de intensidade, por meio de experimentos em várias bases de dados de expressões faciais da literatura.

### **Estudo de caso e definição do *corpus***

O estudo interdisciplinar proposto neste trabalho envolve as áreas de Linguística e Ciência da Computação no que se refere, respectivamente, ao embasamento teórico da análise semiodiscursiva do discurso em programas televisivos e ao desenvolvimento de métodos computacionais para o reconhecimento de emoções dos locutores dos respectivos vídeos. Em especial, este trabalho propõe uma metodologia baseada na técnica de inferência de emoções de Diniz (2013), cuja implementação foi adaptada à luz do mapeamento dos movimentos específicos de expressões faciais proposto por Ekman & Friesen (1978), para identificar e representar os níveis de tensão dos vídeos televisivos, conforme a classificação sugerida no trabalho de David-Silva (2005).

Com o intuito de compreender de forma precisa as dinâmicas de análise discursiva dos vídeos televisivos e dos níveis de tensão intrínsecos em suas unidades constitutivas de informação, detectou-se a necessidade de se realizar este trabalho

utilizando-se, como referencial, um *corpus* constituído por vídeos televisivos da Rede Minas, TV Cultura e Rede Globo, referentes a programas de entrevistas, de debates e telejornalísticos, armazenados em uma base de dados pertencente ao Centro de Apoio a Pesquisas sobre Televisão (CAPTE) do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG). Em grupos disjuntos, esses vídeos foram fontes de pesquisa para os trabalhos de Pereira *et al.* (2015), Braighi-Andrade (2013), Jacob (2013), Conceição (2012), Souza (2012) e Sabino (2011), que integraram os estudos do CAPTE, e dispõem de recursos audiovisuais a serem utilizados nos processos de extração, identificação, classificação e análise dos níveis de tensão.

Na condução da pesquisa em questão, trabalha-se com duas amostras da população de vídeos televisivos adequadas para o estudo dos níveis de tensão em blocos inteiros de um programa, em que cada bloco é um arquivo de vídeo. Como recorte metodológico, essas amostras compõem-se, respectivamente, de 46 blocos dos programas Conexão Roberto D'Ávila, Roda Viva, Brasil das Gerais e Rede Mídia, dos gêneros televisivos debate e entrevista, e de 310 blocos de telejornais que reúnem vídeos do Jornal Minas da Rede Minas, do telejornal MG TV, da Rede Globo, do Jornal da Alterosa, do Jornal Nacional e do Repórter Brasil, exibidos em períodos de gravações de 06 de novembro de 2011 a 30 de julho de 2012.

## **Metodologia**

Esta seção apresenta a metodologia geral usada ao longo do desenvolvimento deste trabalho, incluindo o nível conceitual com a elaboração do modelo proposto e a implementação dos protótipos de processamento do arcabouço. Finalmente, explica a proposta de validação do modelo e do protótipo junto a pesquisadores da Análise de Discurso como estratégia de pesquisa para demonstrar a eficiência e a aplicação do sistema no suporte à análise dos níveis de tensão dos vídeos televisivos.

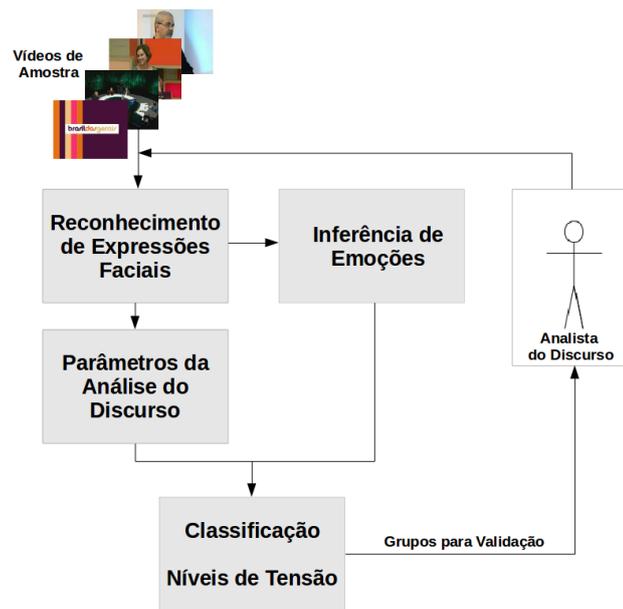
Este artigo trabalha com os recursos obtidos do reconhecimento de expressões faciais para a inferência de emoções em vídeos televisivos, classificando-os em níveis de tensão pré-estabelecidos, conforme as emoções identificadas e devidamente apresentadas em gráficos a serem usados para apoiar as pesquisas em semiótica do discurso desses programas.

Para a etapa de reconhecimento de expressões faciais, este artigo se baseia na detecção de alguns movimentos faciais, conforme o conceito de *Action Units* (AU),

proposto por Ekman & Friesen (1978). Essa abordagem é usada no modelo de inferência de emoções gerado em Diniz (2013), cuja implementação norteia este trabalho, possibilitando adaptações para o cenário televisivo. Os parâmetros discursivos propostos no trabalho de David-Silva (2005) são utilizados para classificar os vídeos nos níveis de *Distensão*, *Tensão Moderada* e *Alta Tensão*, que são explicados a seguir.

Para David-Silva (2005), as exibições televisivas com níveis de *Alta Tensão* (AT) referem-se a assuntos cujo conteúdo remete-nos para uma categoria de conflito, de violência, de tragédia e de morte (homicídios), revelando problemas do mundo que podem produzir uma resposta patêmica no público, principalmente quando os assuntos abordados são mais próximos, social e fisicamente, do público ao qual se endereçam. A *Tensão Moderada* (TM), também denominada por tensão regular, pode ser atribuída a matérias que, mesmo estando no âmbito do conflito, promove uma implicação patêmica menor no telespectador, geralmente quando o palco do fato, os envolvidos ou a temática possuem certa distância do cotidiano do público. Já a *Distensão* (DT) se observa em exibições de programas que nos remetem ao campo semântico da “alegria”, tais como eventos esportivos, comemorações, dicas de culinária, avanços tecnológicos, dentre outros, provocando uma espécie de alívio no telespectador (DAVID-SILVA, 2005).

Sobre esses conceitos, propõe-se classificar nos níveis de *Alta Tensão* aqueles vídeos televisivos em que mais ocorrerem emoções de medo, raiva e tristeza; de *Tensão Moderada* os vídeos que apresentarem, predominantemente, expressões faciais ligadas às emoções de aversão, surpresa e desgosto; e de *Distensão* os vídeos que apresentarem emoções de alegria, recorrentemente, durante a sua exibição.



**Figura 1: Visão geral do trabalho proposto.**

A Figura 1 ilustra uma visão geral da abordagem proposta para a análise automática dos níveis de tensão em vídeos televisivos. Algoritmos para extração de características visuais foram implementados, a fim de fornecer os recursos necessários para a detecção das faces dos indivíduos atuantes nos vídeos. Sobre as faces detectadas, atuam módulos para a extração de expressões faciais. A inferência de emoções é realizada sobre as expressões faciais extraídas. Enfim, as emoções inferidas são apresentadas em gráficos que possibilitam aplicar métricas para o mapeamento dos níveis de tensão dos vídeos e apoiar as pesquisas em semiótica do analista do discurso.

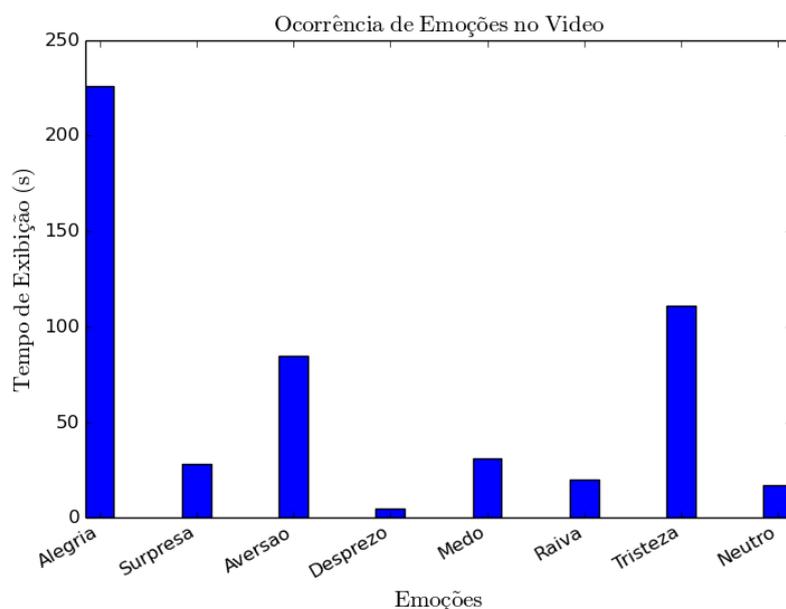
Para a inferência e a visualização gráfica da dispersão de emoções nos vídeos televisivos, este artigo se baseou na proposta de Diniz (2013) que utiliza as AUs detectadas através da leitura do movimento de pontos detectados automaticamente na face identificada por meio de um detector de faces baseado no algoritmo Haartraining de Viola & Jones (2001), alcançando uma acurácia de 85.5%. A categorização das expressões faciais ocorre com a aparição de conjuntos de AUs permanentes (olhos, boca e sobrancelhas) e transientes (rugas causados pelas expressões), que caracterizam uma dentre as seis expressões faciais modeladas. A Tabela 1 mostra o significado e a ocorrência das AUs utilizadas, conforme o modelo de Ekman & Friesen (1978).

Emoção	Ocorrência de AUs e Pontos de Leitura				Representação Visual das AUs Permanentes
	Permanentes	Transientes (Rugas)			
		Geral	Testa	Parte Inferior da Boca	
Alegria	6	-	-	-	Apertar os olhos
	12+25				Boca no formato de sorriso
Surpresa	26	-	-	-	Queda do queixo
	1+2				Levantamento arqueado das sobrancelhas
Aversão	4	Sim	Não	Não	Abaixar as sobrancelhas
	6				Apertar os olhos
Medo	26B	-	Sim	Não	Queda menos acentuada do queixo
Raiva	4	Sim	Não	Sim	Abaixar as sobrancelhas
	10+23+25				Abertura da boca em formato de concha
	18				Espremer os lábios
Tristeza	20	-	Sim	-	Esticar a boca horizontalmente
	4				Abaixar as sobrancelhas

**Tabela 1: Ocorrência de algumas *action units* em expressões faciais.**

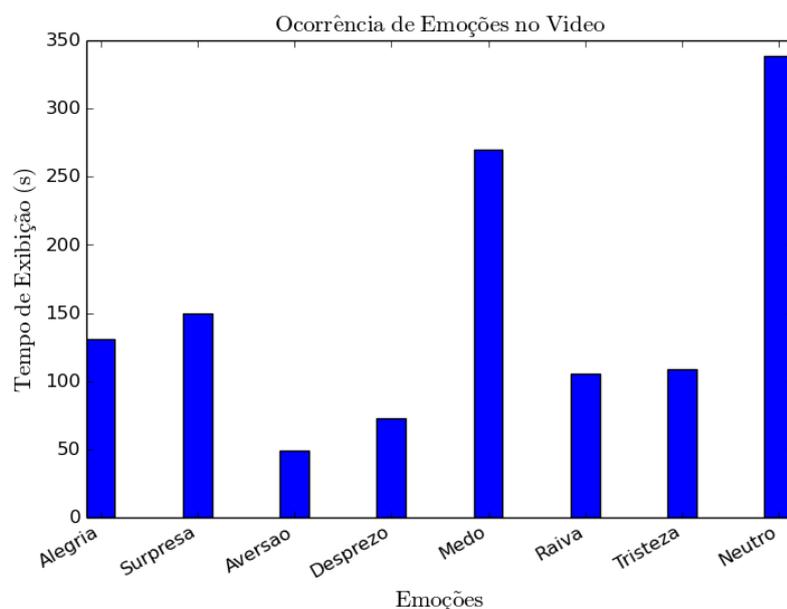
A determinação automática dos pontos de leitura, isto é, pontos de marcação de estruturas de controle como olhos, boca e nariz, ocorre logo após a etapa inicial de detecção automática da face e das regiões de interesse na face. Como o sistema necessita de uma face neutra de expressões faciais para que, sobre ela, os movimentos faciais possam ser lidos sobre os pontos de leitura, definiu-se, neste trabalho, marcar como neutra a primeira face detectada no vídeo televisivo, partindo-se da premissa de imparcialidade em que os apresentadores devem se manter.

A partir do reconhecimento das expressões faciais do vídeo e a inferência de todas as emoções associadas, a próxima etapa é a apresentação desses dados em gráficos para possibilitar a aplicação de alguma métrica e auxiliar as pesquisas em análise dos níveis de tensão do discurso televisivo dos programas. Neste trabalho, propõe-se a visualização dos dados por meio de um histograma, isto é, um modelo de gráfico que retrate a frequência de ocorrência de emoções no vídeo, informando o tempo de exibição, em segundos, em que uma dada emoção foi inferida no vídeo.



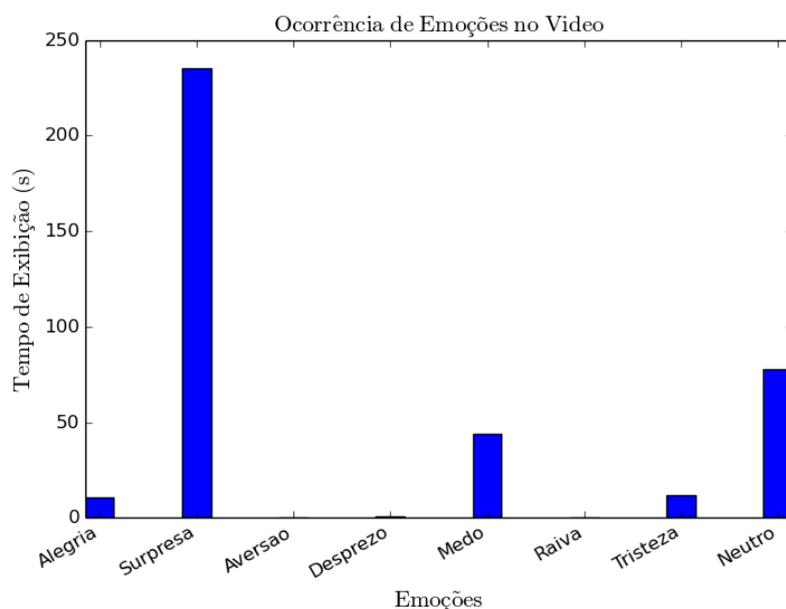
**Figura 2: Gráfico para a ocorrência de emoções no vídeo do programa Conexão Roberto D'Ávila.**

A Figura 2 ilustra um exemplo desse tipo de gráfico para o vídeo referente ao segundo bloco do programa Conexão Roberto D'Ávila que exibiu em 09/07/2010, pela TV Cultura, uma entrevista com o jornalista e escritor Carlos Heitor Cony. O eixo x contempla as emoções modeladas (Alegria, Surpresa, Aversão, Desprezo, Medo, Raiva, Tristeza, Neutro) e no eixo y, as ocorrências de 226, 28, 85, 5, 31, 20, 111 e 17 segundos de cada emoção, respectivamente. Esse exemplo poderia evidenciar uma *Distensão*, visto que ocorreu a emoção de alegria em 226 segundos, porém o tempo de exibição das outras emoções não são desprezíveis e refletem um caso complexo de classificação que evidencia tanto a *Tensão Moderada* (com 118 segundos de exibição para surpresa, aversão e desprezo) quanto a *Alta Tensão* (com 162 segundos detectados para as expressões faciais de medo, raiva e tristeza). Como a categoria *Distensão* foi definida para apenas uma emoção e as outras categorias possuem três emoções, seria prudente realizar uma análise matemática em relação à relevância de cada categoria em relação ao total de cenas que elas representam em termos das emoções reconhecidas.



**Figura 3: Gráfico para a ocorrência de emoções no vídeo do Brasil das Gerais.**

A Figura 3 apresenta o gráfico de ocorrências de emoções, pelo tempo de exibição para o vídeo referente ao primeiro bloco do programa de debate Brasil das Gerais, pela Rede Minas, sobre a questão da violência contra os homossexuais. O eixo x contempla as emoções modeladas (Alegria, Surpresa, Aversão, Desprezo, Medo, Raiva, Tristeza, Neutro) e no eixo y, as ocorrências de 131, 150, 49, 73, 270, 106, 109 e 338 segundos de cada emoção, respectivamente. Existem, assim, 131 segundos referentes a uma *Distensão*, 272 segundos de *Tensão Moderada* e 485 segundos que evidenciarão *Alta Tensão*, dentro dos 20 minutos e 45 segundos da exibição desse bloco do programa.



**Figura 4: Gráfico para a ocorrência de emoções no vídeo do Jornal Minas.**

A Figura 4 acima apresenta o gráfico de ocorrências de emoções pelo tempo de exibição para o vídeo referente ao primeiro bloco do Jornal Minas, exibido pela Rede Minas, sobre o aumento dos casos de violência contra idosos. O eixo x contempla as emoções modeladas (Alegria, Surpresa, Aversão, Desprezo, Medo, Raiva, Tristeza, Neutro) e no eixo y, as ocorrências de 11, 235, 0, 1, 44, 0, 12 e 78 segundos de cada emoção, respectivamente. Existem, assim, 11 segundos referentes à emoção de Alegria (*Distensão*), 236 segundos de *Tensão Moderada* e 56 segundos de *Alta Tensão* em 6 minutos e 12 segundos de toda a exibição desse bloco do programa (312 segundos), configurando uma tendência significativa de ser um vídeo de *Tensão Moderada*.

### **Conclusão**

Este trabalho apresenta uma nova abordagem que integra métodos computacionais para a inferência de emoções sobre expressões faciais a fim de identificar e classificar os níveis de tensão em vídeos televisivos como abordagem complementar e interdisciplinar à análise semi-discursiva sobre programas de televisão.

As expressões faciais são formas de comunicação não verbal amplamente utilizadas em nosso cotidiano. Elas externam nossas manifestações, em relação aos estímulos a que somos submetidos e, dessa forma, são elementos que fazem parte da composição midiática dos programas televisivos, tanto pelas emoções provocadas por

eles, quanto pelas emoções que seus apresentadores expressam, involuntariamente, ou de forma voluntária, se forem pautadas na estratégia comunicativa da rede de televisão. Os movimentos gestuais e a expressividade facial dos locutores, principalmente os apresentadores, podem fornecer indícios sobre a tensão do discurso gerado pelo respectivo programa e o padrão no sequenciamento dos fatos informados nas instâncias de produção desses objetos informacionais.

Neste contexto, este trabalho propõe a leitura dos quadros de vídeos televisivos, como também identificar, nessas imagens, a face de um interlocutor daquele programa, realizando um estudo da movimentação de pontos de leitura da boca, dos olhos, das sobrancelhas e rugas específicas do rosto, ao longo do tempo. Selecionam-se essas características a fim de se obterem expressões faciais que são utilizadas na inferência de emoções, permitindo a classificação dos vídeos em termos de níveis de tensão e a visualização gráfica desses dados, com o objetivo de se oferecer subsídios a pesquisas sobre os programas televisivos, em diferentes áreas, tais como a comunicação, a semiótica e a análise do discurso.

## Referências Bibliográficas

- BETTADAPURA, V. **Face expression recognition and analysis: the state of the art.** ArXiv e-prints, p. 1-27, Março 2009.
- BRAIGHI-ANDRADE, A. A. **Análise de telejornais: um modelo de exame da apresentação e estrutura de noticiários televisivos.** Rio de Janeiro: E-Papers, 2013.
- BUTTURI JR., A.; LENZI, M. C. **Sobre ethos, jornalismo e constituição: o gênero ombudsman.** Revista Working Papers em Lingüística, v. 10, p. 51-67, 2009.
- CHARAUDEAU, P. **Discurso das mídias.** São Paulo: Contexto, 2006.
- CONCEICAO, F. L. A. **Metodologia baseada em mineração de dados para apoio à análise do discurso de telejornais.** Dissertação - Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, Agosto 2013.
- CLARK, D. G. **Radio in presidential campaigns: The early years (1924-1932).** Journal of Broadcasting, v. 6, n. 3, p. 229-238, 1962.
- DAVID-SILVA, G. **A informação televisiva: uma encenação da realidade (Comparação entre Telejornais Brasileiros e Franceses).** Tese - Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

- DINIZ, F. A. **RedFace: um sistema de reconhecimento de expressões faciais para apoiar um ambiente virtual de aprendizagem.** Dissertação - Universidade do Estado do Rio Grande do Norte, Mossoró, 2013.
- FECHINE, Y. **Performance dos apresentadores dos telejornais: a construção do éthos.** Revista FAMECOS - Mídia, Cultura e Tecnologia, v. 1, n. 36, p. 69-76, 2008.
- FRIEDMAN, H. S.; DIMATTEO, M. R.; MERTZ, T. I. **Nonverbal communication on television news: the facial expressions of broadcasters during coverage of a presidential election campaign.** Personality and Social Psychology Bulletin, v. 6, n. 3, p. 427-435, 1980.
- GODOY-COTES, C. S. **O estudo dos gestos vocais e corporais no telejornalismo brasileiro.** Tese - Pontifícia Universidade Católica de São Paulo, Sao Paulo, 2008.
- GOFFMAN, E. The lecture. In: **Forms of talk.** Pennsylvania: University of Pennsylvania Press, p. 162-195, 1981.
- JACOB, H. D. **Desenvolvimento de um modelo de atenção visual para sumarização automática de vídeos de programas televisivos.** Dissertação - Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, Agosto 2013.
- LIAO, C. T.; CHUANG, H. J.; DUAN, C. H.; LAI, S. H. **Learning spatial weighting for facial expression analysis via constrained quadratic programming.** Pattern Recognition Elsevier, v. 46, p. 3103-3116, 2013
- MACGILCHRIST, F. **Discursive tensions in news coverage of Russia.** John Benjamins Publishing Company (Discourse Approaches to Politics, Society and Culture), 2011.
- MADUREIRA, S. Expressividade da fala. In: KYRILLOS, L. (Ed.). **Expressividade da teoria à prática.** Rio de Janeiro: Revinter, p. 16-25, 2004.
- ODIN, Roger. **De la fiction.** Bruxelles: De Boeck & Lancier, 2000.
- PEREIRA, M. H. R.; SOUZA, C. L.; PÁDUA, F. L. C.; DAVID-SILVA, G.; ASSIS, G. T.; PEREIRA, A. C. M. **SAPTE: A multimedia information system to support the discourse analysis and information retrieval of television programs.** Journal Multimedia Tools and Applications, v. 74, n. 2, 2015.
- PIMENTEL, R. M. L. **Memória e apagamento no imaginário dos telejornais.** Londrina: Discursos Fotográficos, v. 5, n. 6, p. 13-33, Junho 2009.
- SABINO, J. L. M. F. A Análise discursiva de entrevistas e debates televisivos como parâmetro para indexação e recuperação de informações em um banco de dados audiovisuais. Dissertação - Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2011.
- SABINO, J. L. F.; DAVID-SILVA, G.; PÁDUA, F. L. C.. **AD e eventos da mídia: uma análise da espetacularização do conflito verbal.** Acta Semiótica et Linguística, v. 19, p. 1-15, 2014.
- SILVA, A. L.. **Ouvir... Ler: uma análise discursiva da estrutura, da apresentação e dos textos em movimento de telejornais de emissoras all news do brasil.** Dissertação - Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2014.

SOUZA, Celso Luiz de. **Recuperação de vídeos baseada em conteúdo em um sistema de informação para apoio à análise do discurso televisivo**. Dissertação - Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2012.

VERON, E . **Il est là, je le vois, il me parle**. *Communications*, Paris: Seuil, n. 38, 1983.

VIOLA, P.; JONES, M. J. **Rapid object detection using a boosted cascade of simple features**. IEEE Computer Society, v. 1, p. 511–518, 2001.