# Toward Interactively Balancing the Screen Time of Actors Based on Observable Phenotypic Traits in Live Telecast

NAIMUL HOQUE, Stony Brook University
NAZMUS SAQUIB, MIT Media Lab
SYED MASUM BILLAH, Pennsylvania State University
KLAUS MUELLER, Stony Brook University

Several prominent studies have shown that the imbalance in the on-screen exposure of observable phenotypic traits like gender and skin-tone in streaming visual media, such as movies, TV shows, and live telecasts can reinforce gender and racial stereotypes in society. To mitigate this problem, researchers and human rights organizations alike have long been calling for more awareness of these issues in the production of visual media. While awareness among media producers is beginning to grow, there are two key technical limitations in current commercial production software that prevent progress toward these goals. First, most commercial production software lack mechanisms that allow a quantification of the screen time of actors based on their observable phenotypic traits, such as gender and skin-tone in the produced video. Second, there are no visual awareness tools that allow producers to balance the exposure of these phenotypic traits during the production of live telecasts.

In this paper, we propose *Screen-balancer*, an interactive visual tool to overcome these limitations. The design of Screen-balancer is informed by a field study conducted in a professional production studio. Screen-balancer analyzes the facial features of the actors to determine phenotypic traits using off-the-shelf facial detection packages, and then facilitates real-time visual feedback to assist media producers in balancing the screen time of these actors in a live telecast. To demonstrate the effectiveness of our approach, we conducted a user study with 20 participants and asked them to compose live telecasts from a set of video streams simulating different camera angles, and featuring a number of male and female actors with different skin-tones. The study revealed that the participants were able to reduce the difference between male and female screen time by 43% and lighter and darker screen time by 44%, thus showing the promise and potential of using such a tool in commercial production systems.

CCS Concepts: • **Human-centered computing → Visual analytics**.

**ACM Reference Format:**
Naimul Hoque, Nazmus Saquib, Syed Masum Billah, and Klaus Mueller. 2018. Toward Interactively Balancing the Screen Time of Actors Based on Observable Phenotypic Traits in Live Telecast. *J. ACM* 37, 4, Article 111 (August 2018), 23 pages. https://doi.org/10.1145/1122445.1122456

## 1 INTRODUCTION

Visual media play an important role in our society—visual content in media can be seen as a reflection of our societal beliefs. Conversely, our societal beliefs can be influenced by media [70]. Therefore, it is imperative that visual media should not propagate gender or racial stereotypes through conscious or unconscious representations of actors and characters. Prior research has

Authors' addresses: Naimul Hoque, Stony Brook University; Nazmus Saquib, MIT Media Lab; Syed Masum Billah, Pennsylvania State University; Klaus Mueller, Stony Brook University.


0004-5411/2018/8-ART111 $15.00
https://doi.org/10.1145/1122445.1122456

shown that the abundant representation of *passive*, *sexual* and *supporting* female characters can increase the likelihood of sexual misconduct in society [35]. Researchers have also shown that if visual content mostly revolves around characters of a specific gender or race, children may learn to think less of marginalized groups in society [11].

One of the primary ways media promote certain gender and racial stereotypes is by allocating imbalanced screen time to actors of different gender and race [86]. Thus, screen time has become an important measure of (mis)representation of different minority groups in visual media [67]. In CSCW 2019, Jang et al. [48] showed that there exists a statistically significant difference between the visual representation of female and male characters in commercial films. Big organizations, such as UNESCO, the Geena Davis Institute, and UN Women have made enormous efforts to bring awareness to these implicit biases and stereotypes through rigorous research [1, 67, 68, 79], in an attempt to make content-creators more conscious to these issues.

However, formulating universal rules to assess screen time in a creative task such as video production is difficult for several reasons. First, content creation is a form of art, often directed by scripts and the artistic view of the creators. For instance, showing more men than women in a male-dominated historical film might not indicate the director is biased towards male characters—it could be a mere requirement for the plot.

Second, the current commercial video production software hardly provide any support to quantify the screen time of the actors based on their observable phenotypic traits such as gender or skin-tone. However, this can be addressed by the recent image analysis and facial recognition technologies, which have become robust, so much so that these technologies are being deployed in many mission-critical applications, such as smartphone authentication [3], airport and home security systems [40, 78]. In fact, the GDIQ tool from Geena Davis Institute utilizes these technologies in quantifying screen-time of different gender in a video [67]. Similarly, Final Draft [25] and WriterDuet [87] tools provide a function to determine whether a written scenario is gender-equitable or not.

Third, while tools like GDIQ can be used to measure the gender representation in post-production videos, to the best of our knowledge, there is no production software offering visual awareness to producers in realtime so that they can balance the exposure of phenotypic traits, such as gender, skin-tone, as a video is being produced.

In this paper, we present ***Screen-balancer*** to address this void. Screen-balancer is an interactive tool that augments the current live telecast systems as follows: it automatically extracts the screen-time of different actors based on two of their phenotypic traits (gender and skin-tone) in pre-production video streams; then visualizes those statistics in realtime, and offers different visual cues so that media producers can balance the screen-time of the actors without compromising their artistic freedom. Screen-balancer seeks to address the following research question (RQ):

***RQ****: Can we incorporate a tool in the current production system to assist media producers to balance the screen time of male and female actors with different skin-tones in realtime, while keeping producers' artistic freedom?*

The design of Screen-balancer is informed by a field study we conducted in a commercial production studio. We observed that during the production of live telecasts, production studio typically uses multiple cameras capturing different perspectives of the scene, such as zooming into a character, or keeping a master angle that shows all of the characters, or only one individual in a frame. Consequently, these camera-feeds are expected to have different distributions of screen time of actors and characters featuring the telecast. We also observed that choosing a camera-feed for telecasting at a given time depends on the producers' discretion. Based on these observations, we designed Screen-balancer to proactively display the gender and skin-tone distributions for each camera-feed using easy-to-compare bar-graphs. We also present the impact of choosing an individual camera-feed in the near future, enabling producers to make an informed decision as they choose the next camera-feed.

In summary, we make the following contributions:

- The design iterations and development of *Screen-balancer*, a prototype system to balance screen time of actors based on phenotypic traits such as gender and skin-tone.
- A user study with 20 participants to evaluate the effectiveness of Screen-balancer in balancing the screen-time in simulated telecasts.
- Feedback from professional video producers and social science researchers regarding the implications, challenges, and potential deployment issues of Screen-balancer.

The remainder of this paper is organized as follows. In Section 2 we describe prior research related to gender and skin-tone bias in visual media, the limitations of the current production software, and the potential of adapting facial recognition technologies in media production. In Section 3, we present a field study that leads to the design of Screen-balancer in Section 4 and implementation in Section 5. Section 6 presents user study design and evaluation of Screen-balancer tool. We present feedback gathered from professional video producers and academician in Section 7 followed by a discussion in Section 8. Finally, conclusions are drawn in Section 9.

## 2 RELATED WORK

The notion that there exists gender and skin-tone bias in visual media has been shown both in quantitative and qualitative manners. In this section, we describe the various studies done to evaluate and determine effect of gender and skin-tone bias in media, as well as movements against such biases organized by different NGOs and organizations.

### 2.1 Gender Bias in Media

A fair share of existing literature on the relationship between gender and media revolves around identifying how gender bias affects societal beliefs [69, 86]. More specifically, researchers have tried to identify the effects of gender bias posed by media on children or teenagers since they are more susceptible to inherit gender stereotypes at an early age [9, 75, 77]. It has been shown that media acts as a crucial medium on how people at an adolescent age learn about the role of gender in society [83]. Gender portrayal in media has also been shown to affect career choices in later stages of life [34]. Other studies reported effects of media on acceptance of violence against women [66], low self-esteem of women [4, 33], sexual socialization among American teens [84], acceptance of different genders, and racially aware TV shows [10, 11]. Several articles analyzed imbalanced representation of male and female screen-time in live telecast, and discussed implications of such misrepresentation [16, 18, 55].

Considering this crucial role of visual media in our lives, *UNESCO* has declared *Gender Sensitive Indicators for Media (GSIM)*, a list of indicators that ensures equality and women's empowerment in media. GSIM states in the report that the balanced representation of men and women in media is one of the primary indicators of gender equality.

*Geena Davis Institute on Gender in Media* introduced *Geena Davis Inclusion Quotient* or *GDIQ* tool [67], the first-ever automated tool to use machine learning techniques to automatically analyze male and female screen-time and speaking-time in any video content. Using the *GDIQ* tool, the institute has analyzed 200 top grossing Hollywood movies spanning from the year 2014 to 2015 and found out that men appear more than female in almost all types of movies, even in movies with female lead characters. Similar kinds of tools and machine learning based methods have also been developed to analyze visual content from other countries such as Bollywood movies [64], and Bangladeshi TV-shows [46]. Recently, Jang et al. [48] discussed in detail the various adverse affects of gender bias in media, their implications, and attempts to quantify such biases. The authors argued why the Bachdel Test, a popular test used as a measure of gender bias in movies, is not sufficient to encapsulate the complex notion of gender bias in visual contents and proposed eight quantitative indices to quantify the gender bias posed by the visual media. They analyzed 40 movies using off-the-shelf computer vision tools to reaffirm the existence of gender bias based on those 8 indices. While such tools are helpful in finding historic gender stereotypes, their intention is to detect biases from already released visual content. Our tool is different from these tools as we intend to help content-creators mitigate gender bias from their product in the production cycle before it is seen by viewers, by balancing the screen time.

## 2.2 Skin-tone and Bias in Media

Intersectionality is a framework that is often used to define a population through different identity such as gender, race, and class [73]. In this paper, we concentrated on balancing screen time of genders and race in live telecast videos. But race is not a observable phenotypic trait and the task of detecting a person's race solely from a face image can be challenging even for humans, especially when one is not familiar with the race of the characters appearing in the video. Neural networks have shown good performance on detecting race from images for people from some specific geographic regions [80] but not in general. We thus opted for skin-tone as a feature instead of race because we wanted our tool to be generally applicable – we cannot guarantee that a user will be accustomed to the races that can appear in the video and so take an informed decision based on racial perception. Conversely, skin-tone is easy to recognize visually for humans and computers alike and there is very little chance of disagreement among different users in this matter. Skin-tone has been used to identify race in many cultures, albeit this is a method that has seen some debate in the literature [17, 52].

As gender bias adversely affects women, discrimination based on skin-tone disadvantages dark-skinned people [39]. This phenomenon is often known as Colorism or Shadeism. The idea behind racism and colorism is quite similar since in both phenomenoa darker skinned people are discriminated on. Ben-Zeev et al. (2014) found that educated African-American men appear lighter in the mind of their peers [7]. Other researchers linked darker skin-tone to smaller incomes, lower marriage rates, longer prison terms, and fewer job prospects [39, 43, 45].

Numerous studies in Colorism and Media found the existence of discrimination against dark-colored people in movies, tv-shows, and news presentation. Travis L Dixon authored a series of publications to prove that media outlets consistently portray black people as poor, violent, and dysfunctional, whereas white people are portrayed as stable and welfare-oriented [22–24]. Several advertisements from companies such as L'Oreal[1], Elle Magazine[2] has been accused and criticized for whitening skin-color of their models. Such a distorted representation creates segregation in the society and places lighter-skinned people as a superior kind.

## 2.3 Video Analysis and Interactions

Screen-balancer is designed to emulate an actual live telecast setup used in TV studios (Section 3). In this section, we discuss the current video analysis research within the CHI and CSCW community.

Research around video analysis is concentrated on Campaign videos [19, 20], Live-streaming [27, 32, 41, 60–62, 65, 76, 81, 85], behaviour analysis in Video Networks [59, 63, 89], Gaming [53, 57],

[1]https://www.theguardian.com/media/2008/aug/08/advertising.usa1
[2]https://www.telegraph.co.uk/news/celebritynews/8005734/Elle-magazine-in-Gabourey-Sidibe-skin-lightening-controversy.html

and Computational Journalism [21]. VidLyz [20] is an interactive visual tool introduced to help novice video makers produce persuasive campaign videos.

Understanding the patterns of live streaming in different social networks have become a popular research topic since the widespread adaptation of social networks and the introduction of fast internet all over the world. For example, Lu et al. [60] discussed how live streamers in China are using different social media to promote practices that are in danger of getting lost, i.e. Intangible Cultural Heritage (ICH). Fraser et al. [32] found that there are four common types of creative livestreams: teaching, making, socializing, and performing. The authors found that audiences become motivated by watching creative livestreams. Faas et al. [27] also showed the effectiveness of creative livestreaming by examining the livestreams of game developement in Twich.tv.

Screen-balancer's multi-stream interface closely resembles the multi-stream interface of Rivulet [41] and Crowdcasting [76]. Rivulet was developed to facilitate multiple video streams of any event in a mobile platform. In Rivulet, users can see multiple streams in the interface and then choose one of them as a focused stream at any moment. Crowdcasting presents a similar prototype to allow users to interact with multiple video streams. Both Rivulet and Crowdcasting were motivated by the increasing popularity of video streaming services in different mobile and social network platforms. Although our tool processes multiple video streams, unlike Rivulet and Crowdcasting, we are interested in the live telecast setup of TV studios, which requires a rigorous understanding of how live-streaming is being facilitated in the existing TV studio setup.

Other relevant works include MovieBarcodes, Microsoft's "Live Video Analytics", and RTFace. MovieBarcodes is one of the earliest visualization tool for movies, which indexes a movie by using the color palette used in that movie [5, 13]. Microsoft's "Live Video Analytics" [3] runs on a cloud platform to facilitate analyzing videos in real-time. RTFace [82] is another live video analysis platform that uses face recognition system to blur peoples' faces in live video.

### 2.4 Facial Recognition and Commercial Video Production Software

Adobe After Effects supports face tracking in videos. Through face tracking, users can apply effects on facial landmarks such as nose, mouth, pupils, etc. Adobe Photoshop uses facial recognition to organize and search images in a catalog. Similarly, Adobe Premiere allows masking and tracking of moving objects in video through facial and object recognition. Final Cut Pro has a facial detection system to detect faces in a video. To the best of our knowledge, none of the current video production software provide any kind of functionality related to gender or skin-color distribution in a video. This is also true for the live telecast or broadcasting software.

## 3 FIELD STUDY

In order to understand how live videos are produced, two of the authors visited a professional production studio. During the tour, one video producer and several technical crews of the studio accompanied the authors. The video producer conducted the tour to explain the full setup and answered questions

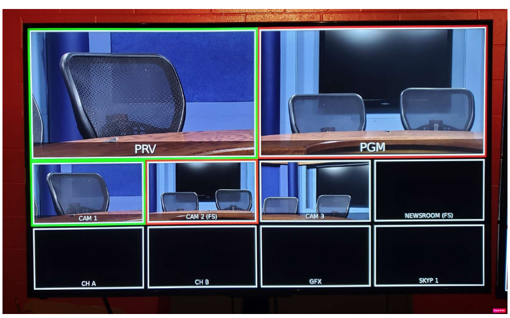
### 3.1 Studio Stage

Figure 1a presents the studio stage at the time of our visit to the studio. The participants of the show perform in this stage and it is what the audience sees on their TVs. This stage is customizable and the decoration varies from show to show. At the time of our visit, the stage had a table discussion setup with three cameras pointing to the table from three directions (figure 1a).

[3]https://www.microsoft.com/en-us/research/project/live-video-analytics/

(a) Studio stage with three cameras

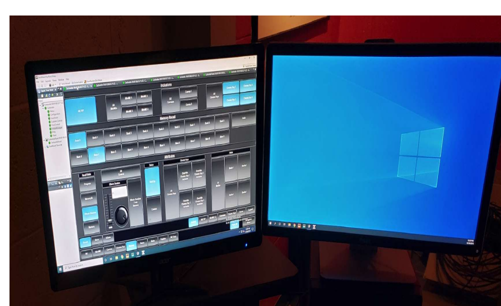
(b) Control Room




(c) Switcher interface: includes three camera feeds from the stage (a), a preview frame (PRV), and a outgoing broadcast (PGM)

(d) Studio Switcher: facilitates buttons to change camera feeds in the Switcher interface (figure c) during a live telecast

(e) Audio Mixer



(f) Controller

Fig. 1. Professional studio setup for live telecast

### 3.2 Control Room

The control room (figure 1b) is the place where the shows are produced and directed. In a usual control room, there are separate panels for processing audio and video. A video producer can see the different feeds coming from the cameras in the staging area (figure 1a) and then select a camera feed for broadcasting, all instantaneously.

*3.2.1 Broadcast Delay.* Even in a live telecast, there is often a 7-10 seconds delay deliberately integrated in the signal path for the purpose of filtering possible bad language and unwanted visual contents. This delay is popularly known as *Broadcast Delay (BD)*. Interestingly, a video producer is oblivious of this filtering task. This censoring is carried out in a separate panel, the Controller (figure 1f). So by the time the feeds are fed into the control room for editing, the feeds seem live to a video producer, even though they are actually delayed by the broadcast delay. This feature is critical in the development of Screen-balancer, as we realized that Broadcast Delay will allow us to analyze gender and skin-tone presence in video frames before they are actually fed into the producer's interface for switching.

*3.2.2 Switcher Interface.* Control rooms are equipped with a visual interface which is used by the producers for selecting camera feeds and seeing what is being broadcast at any moment. The camera feeds or video streams are stacked in the interface (2nd row in figure 1c). The studio that we visited had three camera feeds. namely, CAM1, CAM2, and CAM3. There is a optional *Preview (PRV)* frame in the interface (first rectangle in figure 1c). This frame is used by the producers when they want to see a camera feed in preview before they select the feed for broadcasting. This view is optional and serves the purpose of isolating a camera feed of interest in a bigger screen. The broadcast frame (PGM) is placed right beside the preview frame, although the positions may vary from system to system. We followed this side by side stacked representation of video streams while developing Screen-balancer.

*3.2.3 Switcher.* Figure 1d shows a professional switcher that is used by the producer to interact with the visual interface. Each camera feed has dedicated buttons in the switcher. The producer can see a feed in the preview or select a feed to be broadcast at any moment by clicking on the dedicated buttons.

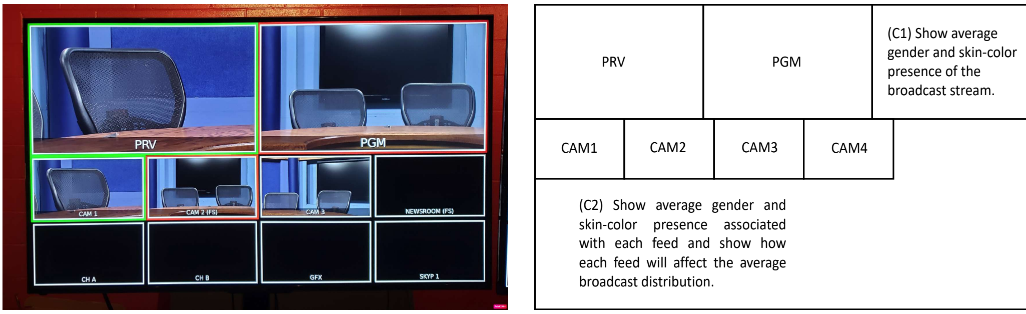
*3.2.4 Audio Mixer.* Figure 1e presents the audio mixer of the studio. At the time of our tour, the mixer was disconnected as there was no ongoing live telecast. During a live telecast, a crew is stationed at the audio mixer for audio production.

*3.2.5 Controller.* The controller is a computer equipped with modern software for editing video and audio. The studio was using a powerful computer with Windows 10 running on it. The computer was equipped with Ross Live Production, a popular software used for managing all the hardware in the studio.

## 4 DESIGN

### 4.1 Guidelines

*"How to enable the screen time balancing functionality in the existing live telecast setup?"*— this is the central challenge that we concentrated on while designing our tool. Our field study enabled us to understand the production setup of live broadcast programs. In particular, we observed that during a live telecast, a video producer constantly changes camera angles to produce the final video. The video producers use a live interactive interface, called *the switcher interface* (figure 1c) to select camera feeds. Interestingly, these camera angles capture different perspective of a live show, thus, have different gender and skin-tone distributions. The current production software do not provide any information on the gender and skin-tone distribution related to each camera feed. The presentation of gender and skin-tone distribution related to each camera angles could allow video producers to take an informed decision in terms of gender and skin-tone screen time while changing camera feeds. Based on that, we focused to provide a solution that can extend the current visual interface and provide additional information of gender and skin-tone distribution. Since the existing setup already provides a visual interface, our solution should also be a visual interface so that the video producers can accustom themselves easily to our system. Our visual tool




(a) Professional Video Switcher Interface

(b) Initial design of Screen-balancer

Fig. 2. Design of Screen-balancer

should resemble the current switcher interface and provide additional information for the purpose of modulating gender and skin-tone presence in a live telecast.

"*How to convey the information related to gender and skin-tone distribution to the video producers?* "— that is the second design challenge we concentrated on. Given the dynamic nature of live telecast and the fact that video producers constantly take quick decisions to produce the video, our system should present these distributions in a way that is easily perceivable by the producers. It is also important to provide a design that is extendable to enable moderation of other observable phenotypic traits. We opted for information visualization for this purpose instead of other means such as text explanation since they provide quick visual perception. Informed by the current live telecast setup, the following design guidelines were developed to integrate the screen-presence balancing process in the *Control Room*.

- G1. The interface would emulate the Switcher functions. Producers would observe the camera feeds or video streams in the interface and be able to interact with the feeds. The interface would closely match the stacked camera feed design of the switcher interface and other relevant multi-stream platforms such as Rivulet [41] and Crowdcasting [76].
- G2. The controller would run available computer vision packages to extract observable facial features from an input video stream. The broadcast delay would allow our system to measure gender and skin-tone presence of frames before they actually reach the switcher interface to be broadcast. After analyzing, the interface should present gender and skin-tone distribution of the live telecast and all the available camera feeds in the interface. This would help users to select the input video stream most suited to balance evolving screen-presence.
- G3. Due to the short broadcast delay, our visual information displays would allow for quick insight and comparisons. Thus, the information should be visualized in a way that is convenient for producers to look over and compare different streams effectively within a short time period.
- G4. The interface should be interactive and should allow easy interaction for changing the camera feeds.

### 4.2 Interface Design

According to G1, our interface should closely emulate the existing switcher interface. We developed the interface iteratively, holding several formal and informal meetings with one video produccer to discuss several aspects of the tool. Figure 2b shows the initial design of Screen-balancer and how it

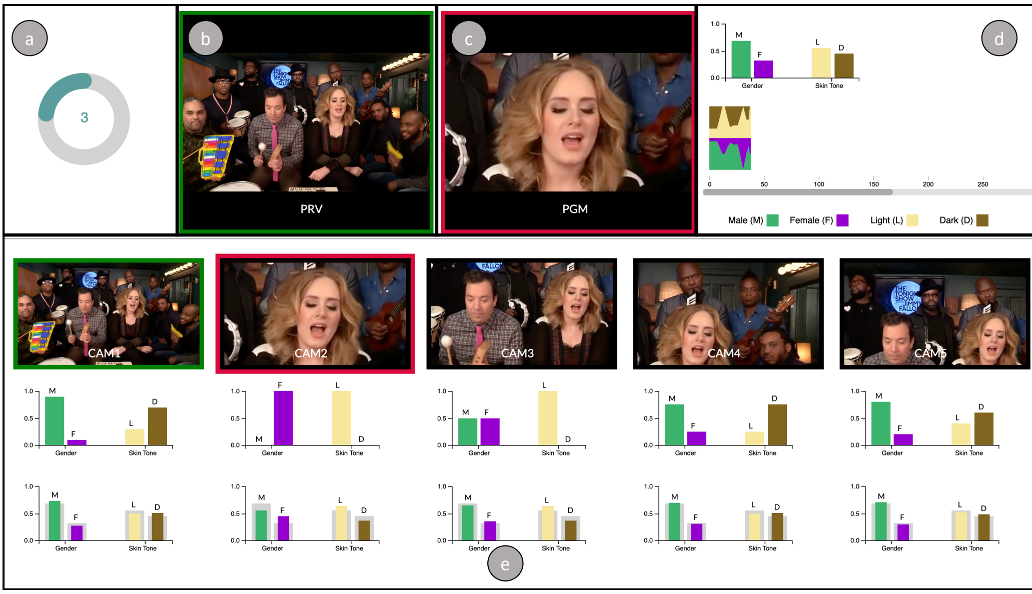



Fig. 3. Components of *Screen-balancer*: **(a)** a countdown timer showing when the charts and graphs will be updated next; **(b)** the preview stream, which video producers use to isolate a particular camera feed from (e) before selecting it for the output stream; **(c)** the output stream; **(d)** bar-charts and a stream graph showing the cumulative distribution and the timeline distribution of sensitive attributes (e.g., gender, skin-tone) in output stream until now; **(e)** input streams (e.g., cameras with different angles and views) along with one bar-chart and one bullet chart each, one showing screen-time of different genders and colors in that stream 10 seconds in advance, and another showing how choosing that particular stream will affect the overall screen-presence distribution of genders and colors in (d) in the next 10 seconds.

was inspired from the current switcher interface. In accordance to G1, our interface has a Preview (PRV) frame, a Broadcast (PGM) frame, and several Camera feeds (CAM). Further, based on G2, we integrated two components to the switcher interface at the first iteration: (C1) visualize statistics related to overall gender and skin-tone portrayal in the broadcast stream; and (C2) visualize statistics related to gender and skin-tone portrayal in each camera feeds and visualize the after-effects of a camera feeds on the overall broadcast distribution.

Figure 3 represents Screen-balancer in a simulated live-stream setup. The visual interface is divided into five different regions. In the next section, we illustrate the design of Screen-balancer by discussing different parts of the interface. Along the illustration, we discuss some of the design choices that were made, how they have evolved during the iterative process, and alternatives that were considered to fulfill G3 in the development of the tool.

*4.2.1 Screen time Distribution (C1).* Region (d) (figure 3) provides two charts (bar chart and stream graph) to present the overall gender and skin-tone distribution of the live telecast. We chose bar charts to represent gender and skin-tone distribution of the broadcast stream and each camera feeds.

These charts update every 10 seconds (see Section 4.2.4). The short duration of this time span requires a data representation that allows users to quickly compare the distributions of each of the camera feeds visually at minimal mental load (according to G3). This type of visual comparison task is often referred as *Comparative Information Visualization* [37, 38]. Bar charts allow easy comparisons based on height and are easy to interpret visually. Pie charts, on the other hand, require users to make comparisons based on angle which can be difficult when the sectors have similar values [30]. Using stacked bar charts in place of a horizontal distribution of bars have similar problems of pie charts. They also make it difficult to compare stack segments with similar values. In their classic graphical perception paper, Cleveland & McGill found that people performed substantially worse on stacked bar charts than on aligned bar charts, and comparisons between adjacent bars were more accurate than stacked bars [15]. We hence concluded that presenting the sensitive variables in the form of simple bar charts, placing the categories such as Male, Female side by side would allow users to quickly evaluate the overall distribution, assess their values visually, and take an informed decision with reduced load.

The color legends for each of the variables are also presented in region d. Here we avoided the stereotyped use of blue and pink for gender data and opted for *green* and *purple* as male and female color respectively, as suggested in [14]. The color green conflicted with the border of the preview frame (PRV), exposing potential confusion on the users part. We evaluated different colors for the preview frame, but the video producer was comfortable with the green color for both the preview frame and the male representation since video producers are familiar with the green border in the preview frame.

*4.2.2 Timeline (C1).* The bottom of Region (d) (figure 3) has the time line view. We use a *(streaming) 100% Stacked Area Chart* because it can visualize the change of skin-tone (top) and gender (bottom) proportions in the video stream over time (stretching along the y-axis). A 100% Stacked Area Chart is essentially a set of stacked and filled line charts, normalized to fit in a box. In order to eliminate small-scale jitter from the display, we applied *Moving Average* filtering to smooth the area chart. That is, for any time $t$, each embedded line graph has a value that is the average of the values from time $t - 3$ to time $t + 3$. The horizontal axis of the area chart dynamically increases as the time increases in the live-stream.

Our initial design did not include this area chart in the interface. It was included in the second iteration since the video producer who we worked with suggested that some sort of historical representation of gender and skin-tone distribution may help them understand the overall trend.

*4.2.3 Input Camera Feeds (C2).* The bottom row of Screen-balancer (region e of figure 3) shows the set of input camera feeds in a horizontal arrangement. At any moment during the live telecast a user can select any of the camera feeds which will connect it to either the preview (PRV) or the the main video (PGM) in regions b or c. As stated in Section 3.2.2, producers often use the PRV to isolate a camera feed from the pool of camera feeds available before selecting it for the actual broadcast. In Screen-balancer, the broadcast feed is highlighted with a red border while the preview feed is highlighted with a green border. To select a camera feed for previewing, a user can either use designated keys assigned for each camera feed or click on the camera feeds. After previewing, a user can select that feed for broadcasting by hitting the "ENTER" button. This mechanism of changing camera feeds is identical to that of a professional video switcher except we use the keyboard to facilitate the functionality of changing camera feeds.

Our interface also allows the switching functionality without the preview option. The transformation is simple, hitting on a designated key, or clicking on a feed will change the broadcast stream with that particular feed directly, without any previewing. To visually convey the flow of video data we draw a red-colored arc between the selected stream and the output stream. Although such colored arcs were appreciated by the producer while using the tool without previewing, they found such connectors distracting while using the previewing functionality. Thus, the connectors are not visible in figure 3, although they are visible in the interaction demonstration (figure 4).

Each camera feed in the region (e) is accompanied by two bar charts. The first chart associated with a feed shows the distributions of gender and skin-tone in the next 10 seconds calculated from that feed.

The second chart associated with each camera feed is a variant of bullet graph [29], which shows the after effect of selecting a camera feed. The bullet graph is known for its usefulness in comparative analysis and can show both reduction and increase of a value in the bar. The bullet graph became popular as a replacement of dashboard gauges and meters, which suffer from perceptual distractions [88]. Bullet graphs can also be used to depict multiple comparisons, percentiles, and other statistical features. These statistical features are omitted in our system as we are interested in the overall distribution of gender and skin-tone. Further, these features can be overwhelming for a video producer, especially with the small amount of time available to glance over all the camera feeds and select a camera feed.

The main objective of this chart is to present how the selection of a feed effects the average distribution in region (d). The broadcast delay allowed us to calculate the gender and skin-tone distributions ahead of the actual streaming. The current value of any category is shown in grey bars in the bullet graph. The effect of choosing a camera feed is shown as colored bars. The *bullet graphs* are included for quick comparisons – at any moment the user needs to only see the colored bar to determine the after effect, and the direction of impact, when choosing a camera feed.

*4.2.4 Countdown Timer.* Camera feeds are dynamic, which suggests the visualizations must be updated dynamically. One of the challenges that we faced while designing Screen-balancer is to figure out a update mechanism that is not overwhelming for a video producer. According to G2, we can analyze frames 7-10s ahead, depending on the broadcast delay. In accordance with the broadcast delay, we can also update the visualizations in every 7-10s. The video producer that we worked with participated in several trials with different number of camera angles and different delays to obtain a comfortable broadcast delay, which is 10s. Region (a) includes a 10 second countdown timer (consistent with the broadcast delay) to indicate when the charts will be updated. The timer resets to 10 upon reaching a value of 0.

Figure 4 presents an example scenario to demonstrate how Screen-balancer can be used to balance visual presence in the video streams.

## 5 DEVELOPMENT OF SCREEN-BALANCER

We used Python as a server side language to process the video streams in our system. A frame rate of 1 fps was chosen to analyze the video frames. The rationale behind this is that the faces in the screen will very rarely change within 1 second.

### 5.1 Face and Gender Detection

Face detection is the most important part of our analysis pipeline as most of our analyses are based on the attributes extracted from faces. We used dlib, an open source Python/C++ library, to detect faces from a video frame.

For the gender detection from face images, we have used a trained convolutional neural network model as proposed by Levi and Hassner [56]. The model is claimed to have a 86% accuracy rate on the Adience benchmark [26]. We created our own test set of 1000 faces extracted from our video sources and found our accuracy rate to be 89%. This test set is also utilized to measure fairness of

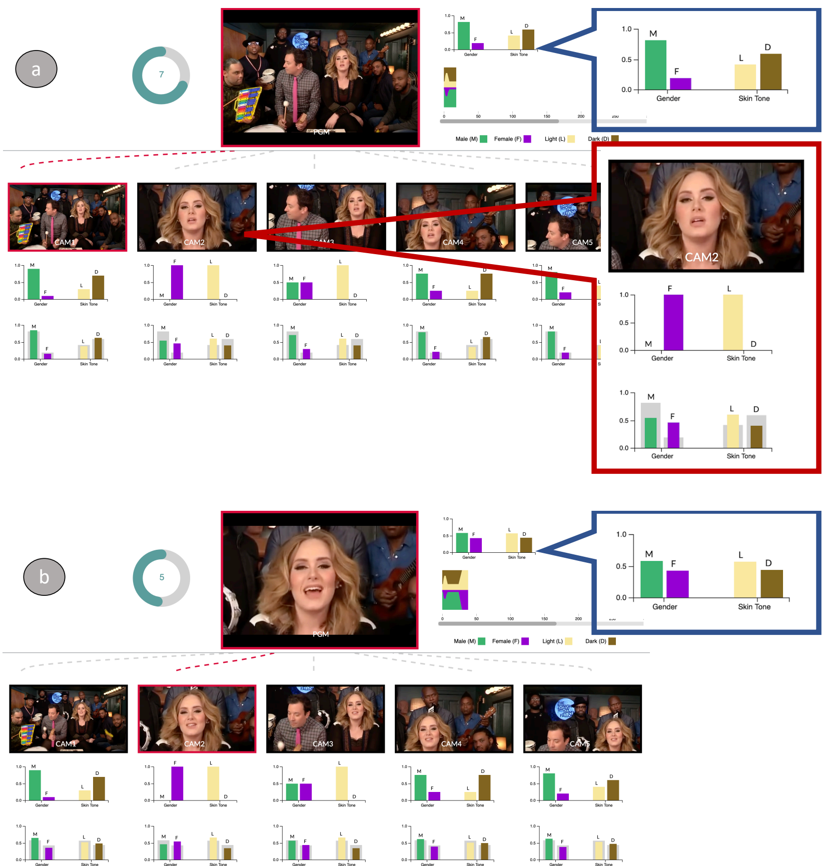



Fig. 4. Interaction with *Screen-balancer* (without the preview). (a) During live-streaming the producer observes that both the gender and skin-tone distributions are unbalanced (blue arrow). The producer decides to change the feed to balance the distribution. By glancing over all the distributions and the imminent effect associated with each camera-feed, the producer decides to select the second camera-feed, as it will balance both the gender and the skin-tone distribution (red arrow). (b) The distribution of gender and skin-tone few seconds later, which are more balanced than (a).

the gender and skin-tone detection models.

Automatic face recognition, and gender recognition are prone to biases since they are often trained using historically biased datasets. For example, recent research has shown that three automatic face recognition systems are more likely to miss-classify darker women than white men or white women [12]. These are valid concerns and it is critical to evaluate the fairness of the dataset and models used in our system. We concentrated on evaluating whether the model detects males and females with similar accuracy or not, quantifying demographic parity. We evaluated the accuracy rate for detecting males and females from the aforementioned curated test set. We found that the model did perform better on detecting male faces (93%) than female faces (87%), exposing potential biased behaviors. The gap remained at 3-4% as we increased the test set size gradually which suggests a moderate disparity. We opted for this model as this is a widely used open-sourced gender detection model and other commercial gender detection models have shown to have disparate impacts [12].

### 5.2 Skin Tone Detection

For detecting skin-tone from face images, we leverage unsupervised learning ($k$-means) for labelling the face images. The method is similar to the one Hoque et al. [46] used to estimate skin-tone in TV serials. Our skin-tone labels are *Light*, and *Dark* in accordance to the skin-tone labels used in [12, 46]. This label is based on the Dermatologist approved Fitzpatrick's six-point skin type scale [31], where Type I, II, and III are labeled as light skin-tone and Type IV, V, and VI are labeled as dark skin-tone [12]. We used the *CelebA* dataset [58] to train our skin-tone detection model. *CelebA* contains 202, 599 face images of 10, 177 different celebrities. For each image we used dlib's landmark detection framework [50] to draw a bounding box around a face excluding the hair area. The pixels around eyes and mouths were also excluded as those pixels might disrupt the skin-tone estimation if the person wears sunglass or shade, or has lipstick in the mouth or has a mustache.

We apply $k$-means on the selected pixels with $k = 2$ as used in [46]. Since, the images only contain face pixels and we have excluded possible disrupting pixels from the faces, the biggest cluster among these 3 clusters should contain the pixels that will indicate the skin-tone of the faces. We take the centroid of the biggest cluster as an estimation of the skin-tone of the faces. After we get the estimations of the skin-tones (hexadecimal values), we again apply $k$-means with $k = 2$ on these skin-tone estimations to find the skin-tone class of each image. Based on the predefined classes and the clustering result, we assign the appropriate skin-tone label (*Light and Dark*) to each image. To evaluate the performance of the labelling task, one of the authors measured the accuracy on a set of one thousand images. The accuracy rate for the labelling task was 98%. Finally, we train this labeled *CelebA* dataset on a two hidden layered Convolutional Neural Network. The network achieved an accuracy rate of 93% on the testing set of *CelebA*.

Similar to the gender detection model, we evaluated the *CelebA* dataset for potential disparate representation. CelebA has shown to have excellent demographic parity, and good equality of opportunity in terms of skin-tone [72]. For evaluating the fairness of the detection model, we calculated the accuracy rate for both lighter and darker skin-tone faces from the curated test set from Section 5.1. The accuracy rates were comparable for both categories (89% and 91% for lighter and darker toned faces respectively).

## 6 EVALUATION

We conducted a user study to assess the effectiveness and usability of Screen-balancer. Specifically, we aimed at validating the following two hypotheses:

- **H1**: A live video production software with Screen-balancer will be more effective in balancing screen-time, than without it.
- **H2**: The Screen-balancer tool will be simple and easy to use.

### 6.1 Participants

We recruited 20 participants (11 males, 9 females) through local mailing lists, social networks, and word-of-mouth. The participants varied in age from 19 to 35 ($M = 25.9$, $SD = 3.81$), gender ($Male = 11$, $Females = 9$). Our inclusion criteria included familiarity with video editing, content creation, post-production, news broadcasting, and live streaming. Among the 20 participants, there were two professional filmmakers, one camera operator, one video advertisement maker, three

Fig. 5. Example of cropping a single video to create multiple camera feeds. The cropped videos on the right were used to simulate a live telecast in our study.

journalists, and the rest were aspiring YouTubers. All of them had undergraduate degrees in different majors.

### 6.2 User Task and Study Setup

The participants were asked to produce a 3-minute video clip from 5 different camera feeds, while balancing the distribution of male and female actors in the final clip, as well as their skin-tones. In a live telecast, it is important to focus on the current event, for example, to focus on the person who is currently speaking in a news debate. Although it was understood by all the participants since all of them had video-making experience, they were instructed to maintain coherency of the live telecast.

To acquire real-world, pre-production videos for our study, we sought to collect videos that were available online and were captured from different camera angles to simulate a live-stream setup. Finding such videos was difficult, because content creators already edited and compiled these videos from different camera angles. We specifically searched for videos that (i) were shot as a single camera angle, or (ii) had multiple angles stacked together in a single frame. We found several musical shows meeting our first criterion; we cropped out different parts of these videos to simulate different camera angles. We also found several discussion shows meeting our second criterion, in which we cropped each portion showing one or more character to re-crate the live studio setting. An example of cropping a single video to yield multiple camera feeds is shown in figure 5. While cropping, we ensured that the gender and skin-tone distribution in each cropped feed was different, and no single feed provided absolute parity to the average distribution of gender and skin-tone. We also ensured that the feeds were dynamic, meaning that the distribution of gender and skin-tone changed over time in the camera feeds. To ensure that the simulation imitated an actual live telecast, the videos and different camera angles were reviewed by a video producers. Based on their feedback, the videos were refined. In total, we prepared 6 videos from 3 different genres, such as talk-shows, news, and live music performance. We provided explanation to the participants that the live telecasts are simulated and the video streams are prepared from an already aired video.

To facilitate remote participation, we deployed our prototype on the web, hosted on a MacBook Pro laptop running an Apache web server. Five participants conducted the experiment remotely, and the rest conducted it on the aforementioned MacBook Pro laptop in our lab. When administrating remote studies, the experimenter communicated over Skype.

### 6.3 Design

We conducted a repeated measures within-subject experiment.

**C1. Simple Video Editor**: This prototype represented a basic video switching software (e.g., figure 1c), where users could choose a camera feed from a list of available feeds at any time, to produce the final live telecast. This was our *baseline*.

**C2. Simple Video Editor + Screen-balancer-Lite**: This prototype only visualized the average and timeline distributions of screen time in the broadcast feed, as shown in figure 3d. This condition is included in the study to understand the effectiveness of the visualization employed in the system in a gradual manner.

**C2. Simple Video Editor + Screen-balancer**: This was the full prototype, as shown in Figure 3.

To minimize the learning effect, we counterbalanced the ordering of study conditions and task videos. We allotted ~20 minutes for each practice session. At the end of each condition, we administrated a NASA-TLX [44] and a SUS questionnaire [8] about the recently used video editor. The experimenter took notes during the session. All sessions were video recorded and transcribed. Each session lasted for 60 minutes, and culminated with participants making suggestions and recommendations.

### 6.4 Data Collection and Analysis

We analyzed the experimenter's notes, logs and recorded videos to measure the following metrics: (i) the distribution of males and females in the produced video clips; (ii) the distribution of skin-tones (e.g., light, and dark); and (iii) the number of times participants changed camera feeds. We also calculated the following subjective measures: (i) coherency of the produced video clips when using Screen-balancer, and (ii) perceived workload and SUS scores.

We performed a repeated-measures ANOVA with the study conditions as the independent variables. For our statistical analyses, we used the Greenhouse-Geisser correction for correcting violations of sphericity, and post-hoc tests using a paired t-test with the Bonnferroni correction.

*6.4.1 Distribution of Gender (H1).* We found a significant effect of study condition on the differences in gender distribution, i.e., gender presence, in the produced video clips, $F(1.392, 26.453) = 6.602, p < .010$. On average, these values were $0.218$ ($SD = 0.112$) for baseline, $0.159$ ($SD = 0.049$) for Screen-balancer-Lite, and $0.123$ ($SD = .066$) for Screen-balancer, as shown in Figure 6. While using Screen-balancer, the participants reduced the gap between male and female screen-time by 43% compared with baseline, and by 27% compared with Screen-balancer-Lite. However, only the former was found to be statistically significant ($p < .032$). Even though using Screen-balancer-Lite yielded a 27% reduction of screen-time between males and females compared to baseline, this reduction was not statistically significant.

*6.4.2 Distribution of Skin-tone (H1).* We also found a significant effect of study condition on the differences in skin-tone distribution, i.e., skin-tone presence, in the produced video clips, $F(2, 38) = 5.683, p < .007$. As shown in Figure 6, the average difference between light and dark toned actors was $0.183$ ($SD = 0.111$) for baseline, $0.145$ ($SD = 0.069$) for Screen-balancer-Lite, and $0.102$ ($SD = .0518$) for Screen-balancer. Between baseline and Screen-balancer, the difference of screen time was reduced by 44%, which was found to be statistically significant ($p < .027$), as expected. No other comparisons were significant.

*6.4.3 Video Coherency (H1).* To assess the quality and coherency, we asked three human evaluators to rate each video produced by using Screen-balancer on a scale of 3 (1=incoherent, 2=partially-coherent, and 3=coherent). The inter-annotator agreement was substantial (Fleiss' $\kappa = 0.63$). The final verdict was made via majority voting. Out of 20 video clips, 16 clips were rated as coherent, 3 as partially-coherent, and only 1 as incoherent.

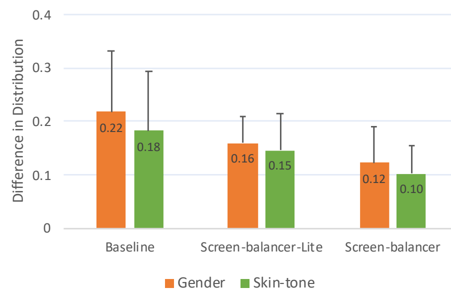



Fig. 6. The average differences in distributions of gender and skin-tones in the final clip produced using 3 study editors. Error-bars show +1 SD.

*6.4.4 Switching Camera (H2).* We anticipated that the participants would frequently switch cameras when using Screen-balancer. Surprisingly, we found that switching camera occurred most frequently when they used baseline ($M$ = 15.250, $SD$ = 8.503), followed by Screen-balancer-Lite ($M$ = 11.550, $SD$ = 5.443), and Screen-balancer ($M$ = 13.650, $SD$ = 7.169). However, these differences were not significant, as reported by a one-way repeated measures ANOVA.

*6.4.5 Perceived Workload and Subjective Feedback (H2).* We found a significant difference among the workload under three conditions, $F(2, 38) = 3.262, p < .049$. The participants reported that balancing screen-time in videos without any aid was mentally demanding ($M$ = 44.542, $SD$ = 16.583). Surprisingly, they reported that their workload decreased the most ($p < .040$) when using Screen-balancer-Lite ($M$ = 35.747, $SD$ = 13.312), rather than using Screen-balancer ($M$ = 37.272, $SD$ = 17.269).

We observed similar trends in the SUS scores – the SUS score increased the most in Screen-balancer-Lite ($M$ = 85.416, $SD$ = 5.103), followed by Screen-balancer ($M$ = 77.083, $SD$ = 5.685), and baseline ($M$ = 51.250, $SD$ = 17.084).

## 6.5 Discussion

The user study results validated our hypotheses H1 and H2. The participants were able to decrease the difference between male and female screen time, and the screen time of actors with light and dark skin-tone. The results were statistically significant between two of the study conditions: baseline switcher, and Screen-balancer. The participants were also able to decrease the difference when using Screen-balancer-Lite with limited visual information, compared to the baseline, but the results were not statistically significant.

We measured cognitive load of the participants through NASA-TLX and SUS-score. Although, the results suggest a minor increase of workload while using Screen-balancer compared to the baseline, the results were not statistically significant. In the post-study discussion, the participants mentioned that glancing over multiple bar-chars at once was the main reason why their workload marginally increased and the SUS score marginally decreased with Screen-balancer as opposed to Screen-balancer-Lite. However, they also mentioned that they would get used to it with practice.

Several participants mentioned that while using the baseline, they tried to remember what camera-feed they had chosen in the past. But at some point they became clueless and started to choose a camera-feed randomly and more frequently, which was also evident in the switching statistics in Section 6.4.4.

The participants were also instructed to maintain the coherency of the show. Three human evaluators audited the videos afterwards and found only one as being incoherent. One of our evaluators who was a professional film-maker made the following comment: *It is very difficult to measure how good a video is. In terms of coherency, it is people's natural instinct to go with a video that has activity in it. Having said that, you often see live shows that frames one person and someone else is actually talking in the show. No one complains about them. Yes, I saw some minor inconstancy in the videos, but that is probably because the gender and skin-tone distributions among the videos were itself disproportional which is also true for many real-life tv shows.*

Despite promising results of our study, the tool still needs to be evaluated by live video producers. Although all of our participants were video-makers and had video processing and editing experience to some extent, feedback from live video producers would enable us to understand some of the issues that came to our knowledge, such as video coherency, and cognitive load. Motivated by that, we interviewed three professional live video producers and two researchers in the evaluation of Screen-balancer which follows next.

## 7 INTERVIEW SESSION WITH PROFESSIONALS

After the development of the tool and the user study, we interviewed three professional video producers who were not involved in the development of Screen-balancer, and two academician working in the field of gender studies and diversity inclusion. We will refer to the three video producers as VP1, VP2, and VP3 and the two academician as AC1, and AC2. Among the interviewees, two of them were female (VP1, AC1) and the others were male (VP2, VP3, and AC2). Two of our interviewees had a dark skin-tone (VP2, and AC1), while the others had a lighter skin-tone. The interview sessions started with a brief explanation of the tool, followed by a demonstration of the tool, and ended with a semi-structured interview. We sorted the feedback into the following five thematic categories:

### 7.1 Resemblance to Video Switcher

All three video producers were comfortable with the complexity of our system. To that end, we did not have to explain the interaction techniques used in our system as they were identical to those of a professional switcher, and all of the producers were already trained to work with these types of systems. The first impression of VP2 was "*Ohh! that is a switcher!*". The bar charts employed in the system were also easily grasped by the producers, although they did not understand at first the bar charts that depicted the after effects of selecting a feed.

When asked about the complexity of the system, VP2 replied, "*My mind is trained to operate in a manner that allow me to glance multiple feeds simultaneously and act dynamically to produce a video. Your system is quite similar to what I am used to, so, no, I do not think it is a complex system.*" VP3, on the other hand, suggested a training session for a new user before an actual live telecast.

### 7.2 "Exciting" and "Timely Solution"

All three video producers expressed their admiration toward the automated nature of our tool. They tagged our tool as "exciting" and "a timely solution". Two of the video producers mentioned that they were very cautious about how they portray different groups in a show, and that our tool can help them achieve this goal more effectively.

During our interview session we came to know that live telecasts such as news presentations are often scripted. In VP1's opinion, Screen-balancer may not be useful in such cases due to the lack of uncertainty, but she said that it definitely had application in a dynamic live telecast.

### 7.3 Potential of Real-life Deployments

VP2, who is associated with a commercial production software, was interested to know technical difficulties associated with automated gender and skin-tone detection. He was excited by the prospect of such a system integrated in his own product. In his own words, "*Commercial production software are always competing with each other and looking for new features to integrate in their system. I believe, they would be very excited at the prospect of this tool.* "

He suggested to engage into a partnership with a professional studio or research institute to validate the tool in an actual live telecast setup, which would bolster our claims.

### 7.4 Improvements

The video producers mentioned several other aspects of video production which they thought if automated would decrease their cognitive load. For example, they mentioned that they often have a high level idea of how much screen-presence to allocate for each characters, but then find it tedious to track each character in the actual live telecast. Indeed, it is possible to track faces in video production software such as Adobe After Effects, but no live telecast software provides such functionality. Our tool presents an interface design that can be extended to include other visually observable or detectable features such as postures or gesture. This information could be added to the interface to provide assistance to the video producers. VP1 suggested that the integration of more features into the system would make the system more "lucrative", and would "motivate" video producers to use the tool.

### 7.5 Gender and Skin-tone Bias in Media

All the interviewees associated our work with gender and racial bias, even without any explicit claim from us. All of them agreed that imbalanced representation of different minority groups is problematic and they shared their own professional experience on the subject matter. AC1, and AC2 also shared several media portrayal stereotypes that they had studied over the years and tagged our research as an "Action Plan" against media stereotypes and biases.

AC1 had some concerns about the potential misuse of our tool. He asked *"what if someone uses Screen-balancer to increase bias in the produced video"?* There is no straightforward answer to this question, but probably the best answer is to increase the consciousness among content creators. Some other participants also mentioned that there is a certain danger that people might "game" a system such as this when it becomes too transparent. In the literature this is known as *transparency-gameability trade-off*. This is a well-known phenomenon and as discussed by Ghani [36], and Rudin [71] the phenomenon is often used as an excuse to make opaque systems. Transparency could motivate producers to create balanced videos, whereby attempting to game it would reveal unwillingness to portray better representation of minority groups.

## 8 DISCUSSION AND LIMITATIONS

In this section, we discuss various aspects that came to our attention during the user study, the post-study interview, and the demo and interview session with professionals.

### 8.1 RQ: Gender and Skin-tone Moderation in Live Telecast

Screen-balancer is the first-ever design to enable a screen-balancing task integrated in an already existing live telecast setup. Our tool is powered by Facial Recognition that allows automatic extraction of gender and skin-tone presence. Our user study showed that users were able to decrease the screen-time gap between different genders and skin-colors significantly. Further, feedback from live video producers revealed that Screen-balancer is easy to use and has application in an actual live telecast setup which proved our RQ: Gender and skin-tone moderation can be integrated in the existing live telecast setup and can help video producers to balance visual presence of gender and skin-tones in live telecast.

There were some concerns about the accuracy of our detection system. One video producer asked *"Can the system detect a female with short hair"?* The deep networks employed in our system are robust to these types of visual distraction, but we can not guarantee that the system is always accurate. For example, the lighting of a scene can affect skin-tone of a characters appearing in the video. A person with lighter skin-tone may appear darker to our detection system in absence of adequate light. Although we are interested in determining traits that are observable in the video, and we explained this to our study participants and interviewees, a video producer may disagree with the system in these cases. This raised a limitation in our system: "What happens if a video producer and Screen-balancer disagrees on someones gender or skin-tone?" At this moment, Screen-balancer does not have any mechanism to handle this situation. One way to mitigate this problem is to integrate a visual correctional system in the interface which the video producers can control. The detection system can then be refined using Active Learning [74] or One-shot Learning [28]. There were also concerns about whether our system is capable of detecting transgender or not. As discussed by [51], most automatic gender recognition software systems do not recognize transgender. Finally, there is the possibility of the models being biased, as discussed in several research articles [2, 6, 42, 47, 49]. Although we evaluated the fairness of the models employed in our system, the analysis is not extensive and we can not guarantee that our models are fully fair since the definition and measurement of fairness varies from system to system [54].

All these concerns reveal the shortcomings of facial recognition which are widely recognized by the machine learning and computer vision community and are active research topics. Despite these shortcomings, facial recognition based systems are employed in a wide range of applications. Our method advances the research on facial recognition assisted video production tools. Any fair and accurate model can be integrated in our system to detect faces, genders, or skin-tones.

### 8.2 Issues with Video Coherency and Cognitive Load

To understand the cognitive load associated with the Screen-balancer, it is important to understand the cognitive load of a video producer who is planning and scripting a live telecast, dynamically changing camera feeds to create a coherent live show. None of the video producers we interviewed found our tool overwhelming. While it is understandable that a users might need significant training to adapt to the visualization employed in our tool, we note that quite a few of our study participants were unfamiliar with professional switching activities and this may have been the reason behind the concerns about the video coherency and the cognitive load.

The notion of video coherency is difficult to measure quantitatively. The term coherency can mean different things for different persons as video production depends mostly on the producers and directors personal choices. One possible solution that emerged from our discussions is to increase the broadcast delay. This could make our interface less overwhelming for the content creators and would allow them to concentrate more on the artistic aspects. Our current system provides a broadcast delay of 10 seconds. Although a 7 to 10 seconds delay is usual, longer delays lasting for hours are also common. For example, in the US, cable TV often delays a show filmed on the East Coast by up to three hours for airing it on the West Coast.

Screen-balancer may also be updated to include some video coherency measures in the future. Automated systems can be used to identify body postures, and gestures such as talking, dancing, playing guitar, etc. from a video to indicate which camera angles have activities in the delay cycle. These types of systems will provide guidance for video-creators to select camera angles based on future activities, gender, and skin-tone distribution. The effectiveness of these types of system is subject to further research.

## 9 CONCLUSION

We describe Screen-balancer – a facial recognition based interactive visual analytics tool that assists content moderators or producers to balance phenotypic human traits, such as gender and race, in a live telecast. Our system can be extended and integrated into any production system, and it can also be generalized to moderate other types of visual characteristics such as hair color. Our methodology tackles a difficult problem, and to the best of our knowledge there is no live production system that provides feedback on visual characteristics as the video is being produced. In other words, important issues in diversity such as screen time of different gender and skin-tone has not yet been acknowledged to be moderated by the production systems in use. Screen-balancer can bridge this gap.

## REFERENCES

[1] 2019. UN Women. https://www.unwomen.org/en Accessed: 2019-04-03.
[2] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 199.
[3] Apple. 2018. *Use Face ID on your iPhone or iPad Pro.* https://support.apple.com/en-us/HT208109
[4] Jennifer Stevens Aubrey. 2006. Exposure to sexually objectifying media and body self-perceptions among college women: An examination of the selective exposure hypothesis and the role of moderating variables. *Sex Roles* 55, 3-4 (2006), 159–172.
[5] Mauro Barbieri, Gerhard Mekenkamp, Marco Ceccarelli, and Jan Nesvadba. 2001. The Color Browser: A Content Driven Linear Video Browsing Tool. In *ICME*. Citeseer.
[6] Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2017. Fairness in machine learning. *NIPS Tutorial* (2017).
[7] Avi Ben-Zeev, Tara C. Dennehy, Robin I. Goodrich, Branden S. Kolarik, and Mark W. Geisler. 2014. When an "Educated" Black Man Becomes Lighter in the Mind's Eye: Evidence for a Skin Tone Memory Bias. *SAGE Open* 4, 1 (2014), 2158244013516770. https://doi.org/10.1177/2158244013516770 arXiv:https://doi.org/10.1177/2158244013516770
[8] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
[9] Jane D Brown, Carolyn Tucker Halpern, and Kelly Ladin L'Engle. 2005. Mass media as a sexual super peer for early maturing girls. *Journal of adolescent Health* 36, 5 (2005), 420–427.
[10] Jane D Brown, Kelly Ladin L'Engle, Carol J Pardun, Guang Guo, Kristin Kenneavy, and Christine Jackson. 2006. Sexy media matter: exposure to sexual content in music, movies, television, and magazines predicts black and white adolescents' sexual behavior. *Pediatrics* 117, 4 (2006), 1018–1027.
[11] Jane D Brown and Carol J Pardun. 2004. Little in common: Racial and gender differences in adolescents' television diets. *Journal of Broadcasting & Electronic Media* 48, 2 (2004), 266–278.
[12] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. 77–91.
[13] Manuel Burghardt, Michael Kao, and Christian Wolff. 2016. Beyond shot lengths–using language data and color information as additional parameters for quantitative movie analysis. (2016).
[14] Lisa Charlotte. 2018. *An alternative to pink & blue: Colors for gender data.* https://blog.datawrapper.de/gendercolor/
[15] William S Cleveland and Robert McGill. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association* 79, 387 (1984), 531–554.
[16] Cheryl Cooky, Michael A Messner, and Michela Musto. 2015. "It's dude time!" A quarter century of excluding women's sports in televised news and highlight shows. *Communication & Sport* 3, 3 (2015), 261–287.
[17] Nicholas G Crawford, Derek E Kelly, Matthew EB Hansen, Marcia H Beltrame, Shaohua Fan, Shanna L Bowman, Ethan Jewett, Alessia Ranciaro, Simon Thompson, Yancy Lo, et al. 2017. Loci associated with skin pigmentation identified in African populations. *Science* 358, 6365 (2017), eaan8433.
[18] Kelly K Davis and CA Tuggle. 2012. A gender analysis of NBC's coverage of the 2008 summer Olympics. *Electronic News* 6, 2 (2012), 51–66.
[19] Sanorita Dey, Brittany Duff, Karrie Karahalios, and Wai-Tat Fu. 2017. The art and science of persuasion: not all crowdfunding campaign videos are the same. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 755–769.
[20] Sanorita Dey, Brittany RL Duff, Wai-Tat Fu, and Karrie Karahalios. 2019. VidLyz: An Interactive Approach to Assist Novice Entrepreneurs in Making Persuasive Campaign Videos. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 43.

[21] Nicholas Diakopoulos, Sergio Goldenberg, and Irfan Essa. 2009. Videolyzer: quality analysis of online informational video for bloggers and journalists. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 799–808.

[22] Travis L Dixon. 2017. A dangerous distortion of our families: Representations of families, by race, in news and opinion media. *Color of Change and Family Story research report, December. https://colorofchange. org/dangerousdistortion* (2017).

[23] Travis L Dixon. 2017. Good guys are still always in white? Positive change and continued misrepresentation of race and crime on local television news. *Communication Research* 44, 6 (2017), 775–792.

[24] Travis L Dixon, Erica Bauer, and Christopher Josey. 2018. How News Frames Prime Non-Whites as Social Problems. In *Doing News Framing Analysis II*. Routledge, 343–361.

[25] Final Draft. 2019. *Inclusivity Analysis | Final Draft.* https://www.finaldraft.com/learn/inclusivity-analysis/

[26] Eran Eidinger, Roee Enbar, and Tal Hassner. 2014. Age and gender estimation of unfiltered faces. *IEEE Transactions on Information Forensics and Security* 9, 12 (2014), 2170–2179.

[27] Travis Faas, Lynn Dombrowski, Alyson Young, and Andrew D Miller. 2018. Watch me code: Programming mentorship communities on twitch. tv. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 50.

[28] Li Fei-Fei, Rob Fergus, and Pietro Perona. 2006. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence* 28, 4 (2006), 594–611.

[29] S Few. 2010. Bullet graph design specification. *Perceptual Edge-White Paper* 5 (2010).

[30] Stephen Few and Perceptual Edge. 2007. Save the pies for dessert. *Visual Business Intelligence Newsletter* (2007), 1–14.

[31] Thomas B Fitzpatrick. 1988. The validity and practicality of sun-reactive skin types I through VI. *Archives of dermatology* 124, 6 (1988), 869–871.

[32] C Ailie Fraser, Joy O Kim, Alison Thornsberry, Scott Klemmer, and Mira Dontcheva. 2019. Sharing the Studio: How Creative Livestreaming can Inspire, Educate, and Engage. In *Proceedings of the 2019 on Creativity and Cognition*. ACM, 144–155.

[33] BL Fredrickson and T Roberts. 1997. Toward understanding women's lived experiences and mental health risks. *Psychology of Women Quarterly* 21, 2 (1997), 173–206.

[34] Reuma Gadassi and Itamar Gati. 2009. The effect of gender stereotypes on explicit and implicit career preferences. *The Counseling Psychologist* 37, 6 (2009), 902–922.

[35] Silvia Galdi, Anne Maass, and Mara Cadinu. 2014. Objectifying media: Their effect on gender role norms and sexual harassment of women. *Psychology of Women Quarterly* 38, 3 (2014), 398–413.

[36] Rayid Ghani. 2019. *You Say You Want Transparency and Interpretability?* http://www.rayidghani.com/you-say-you-want-transparency-and-interpretability

[37] Michael Gleicher. 2017. Considerations for visualizing comparison. *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 413–423.

[38] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D Hansen, and Jonathan C Roberts. 2011. Visual comparison for information visualization. *Information Visualization* 10, 4 (2011), 289–309.

[39] Arthur H Goldsmith, Darrick Hamilton, and William Darity Jr. 2006. Shades of discrimination: Skin tone and wages. *American Economic Review* 96, 2 (2006), 242–245.

[40] Google. 2019. *Learn about familiar face detection and how to manage your library.* https://support.google.com/googlenest/answer/9268625

[41] William A Hamilton, John Tang, Gina Venolia, Kori Inkpen, Jakob Zillner, and Derek Huang. 2016. Rivulet: Exploring participation in live events through multi-stream experiences. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*. ACM, 31–42.

[42] Moritz Hardt, Eric Price, Nati Srebro, et al. 2016. Equality of opportunity in supervised learning. In *Advances in neural information processing systems*. 3315–3323.

[43] Matthew S Harrison and Kecia M Thomas. 2009. The hidden prejudice in selection: A research investigation on skin color bias. *Journal of Applied Social Psychology* 39, 1 (2009), 134–168.

[44] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage Publications Sage CA: Los Angeles, CA, 904–908.

[45] Joni Hersch. 2011. The persistence of skin color discrimination for immigrants. *Social Science Research* 40, 5 (2011), 1337–1349.

[46] Md Hoque, Rawshan E Fatima, Manash Kumar Mandal, Nazmus Saquib, et al. 2017. Evaluating gender portrayal in Bangladeshi TV. *arXiv preprint arXiv:1711.09728* (2017).

[47] Jevan A Hutson, Jessie G Taft, Solon Barocas, and Karen Levy. 2018. Debiasing desire: addressing bias & discrimination on intimate platforms. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 73.

[48] Ji Yoon Jang, Sangyoon Lee, and Byungjoo Lee. 2019. Quantification of Gender Representation Bias in Commercial Films based on Image Analysis. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 198.

[49] Matthew Joseph, Michael Kearns, Jamie H Morgenstern, and Aaron Roth. 2016. Fairness in learning: Classic and contextual bandits. In *Advances in Neural Information Processing Systems*. 325–333.
[50] Vahid Kazemi and Josephine Sullivan. 2014. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1867–1874.
[51] Os Keyes. 2018. The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 88.
[52] Razib Khan. 2019. *Skin color is not race.* https://www.discovermagazine.com/the-sciences/skin-color-is-not-race
[53] Soomin Kim, Gyuho Lee, Seo-young Lee, Sanghyuk Lee, and Joonhwan Lee. 2019. Game or Live Streaming?: Motivation and Social Experience in Live Mobile Quiz Shows. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, 87–98.
[54] Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. 2017. Counterfactual fairness. In *Advances in Neural Information Processing Systems*. 4066–4076.
[55] Martha M Lauzen. 2017. Boxed in 2016-17: Women on screen and behind the scenes in television. *Center for the Study of Women in TV & Film. Disponível em: https://womenintvfilm. sdsu. edu/wp-content/uploads/2017/09/2016-17_Boxed_In_Report. pdf. Acesso em* 2 (2017).
[56] Gil Levi and Tal Hassner. 2015. Age and Gender Classification Using Convolutional Neural Networks. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops*. http://www.openu.ac.il/home/hassner/projects/cnn_agegender
[57] Yao Li, Yubo Kou, Je Seok Lee, and Alfred Kobsa. 2018. Tell Me Before You Stream Me: Managing Information Disclosure in Video Game Live Streaming. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 107.
[58] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. 2015. Deep Learning Face Attributes in the Wild. In *Proceedings of International Conference on Computer Vision (ICCV)*.
[59] Xing Lu and Zhicong Lu. 2019. Fifteen Seconds of Fame: A Qualitative Study of Douyin, A Short Video Sharing Mobile Application in China. In *International Conference on Human-Computer Interaction*. Springer, 233–244.
[60] Zhicong Lu, Michelle Annett, Mingming Fan, and Daniel Wigdor. 2019. I feel it is my responsibility to stream: Streaming and Engaging with Intangible Cultural Heritage through Livestreaming. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 229.
[61] Zhicong Lu, Michelle Annett, and Daniel Wigdor. 2019. Vicariously Experiencing it all without Going Outside: A Study of Outdoor Livestreaming in China. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 25.
[62] Zhicong Lu, Seongkook Heo, and Daniel J Wigdor. 2018. StreamWiki: Enabling Viewers of Knowledge Sharing Live Streams to Collaboratively Generate Archival Documentation for Effective In-Stream and Post Hoc Learning. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 112.
[63] Xiaojuan Ma and Nan Cao. 2017. Video-based evanescent, anonymous, asynchronous social interaction: Motivation and adaption to medium. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 770–782.
[64] Nishtha Madaan, Sameep Mehta, Taneea Agrawaal, Vrinda Malhotra, Aditi Aggarwal, Yatin Gupta, and Mayank Saxena. 2018. Analyze, detect and remove gender stereotyping from bollywood movies. In *Conference on Fairness, Accountability and Transparency*. 92–105.
[65] Misa Maruyama, Scott P Robertson, Sara Douglas, Roxanne Raine, and Bryan Semaan. 2017. Social watching a civic broadcast: Understanding the effects of positive feedback and other users' opinions. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 794–807.
[66] Sarah K Murnen, Carrie Wright, and Gretchen Kaluzny. 2002. If "boys will be boys," then girls will be victims? A meta-analytic review of the research that relates masculine ideology to sexual aggression. *Sex roles* 46, 11-12 (2002), 359–375.
[67] Geena Davis Institue on Gender in Media. 2015. *The Reel Truth: Women Aren't Seen or Heard.* https://seejane.org/research-informs-empowers/data/
[68] Geena Davis Institue on Gender in Media. 2017. *The Geena Benchmark Report: 2007-2017.* https://seejane.org/wp-content/uploads/geena-benchmark-report-2007-2017-2-12-19.pdf
[69] Karen Ross. 2010. *Gendered media: Women, men, and identity politics.* Rowman & Littlefield.
[70] Tara Ross. 2019. Media and Stereotypes. *The Palgrave Handbook of Ethnicity* (2019), 1–17.
[71] Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1, 5 (2019), 206–215.
[72] Prasanna Sattigeri, Samuel C Hoffman, Vijil Chenthamarakshan, and Kush R Varshney. 2018. Fairness gan. *arXiv preprint arXiv:1805.09910* (2018).
[73] Ari Schlesinger, W Keith Edwards, and Rebecca E Grinter. 2017. Intersectional HCI: Engaging identity through gender, race, and class. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5412–5427.

[74] Burr Settles. 2009. *Active learning literature survey.* Technical Report. University of Wisconsin-Madison Department of Computer Sciences.

[75] MJ Sutton, Jonathon D Brown, KM Wilson, and JD Klein. 2002. 'Shaking the Tree of Knowledge for Forbidden Fruit. *Sexual teens, sexual media: Investigating media's influence on adolescent sexuality* (2002), 25–55.

[76] John Tang, Gina Venolia, Kori Inkpen, Charles Parker, Robert Gruen, and Alicia Pelton. 2017. Crowdcasting: Remotely Participating in Live Events Through Multiple Live Streams. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 98.

[77] Tom FM Ter Bogt, Rutger CME Engels, Sanne Bogers, and Monique Kloosterman. 2010. "Shake it baby, shake it": Media preferences, sexual attitudes and gender stereotypes among adolescents. *Sex roles* 63, 11-12 (2010), 844–859.

[78] TSA. 2018. *Biometrics Technology | Transportation Security Administration.* https://www.tsa.gov/biometrics-technology

[79] UNESCO. 2019. Gender-Sensitive Indicators for Media. http://www.unesco.org/new/en/communication-and-information/crosscutting-priorities/gender-and-media/gender-sensitive-indicators-for-media Accessed: 2019-04-03.

[80] Thanh Vo, Trang Nguyen, and C Le. 2018. Race recognition using deep convolutional neural networks. *Symmetry* 10, 11 (2018), 564.

[81] Dennis Wang, Yi-Chieh Lee, and Wai-Tat Fu. 2019. " I Love the Feeling of Being on Stage, but I Become Greedy": Exploring the Impact of Monetary Incentives on Live Streamers' Social Interactions and Streaming Content. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 92.

[82] Junjue Wang, Brandon Amos, Anupam Das, Padmanabhan Pillai, Norman Sadeh, and Mahadev Satyanarayanan. 2017. A scalable and privacy-aware IoT service for live video analytics. In *Proceedings of the 8th ACM on Multimedia Systems Conference*. ACM, 38–49.

[83] L Monique Ward. 2002. Does television exposure affect emerging adults' attitudes and assumptions about sexual relationships? Correlational and experimental confirmation. *Journal of youth and adolescence* 31, 1 (2002), 1–15.

[84] L Monique Ward. 2003. Understanding the role of entertainment media in the sexual socialization of American youth: A review of empirical research. *Developmental review* 23, 3 (2003), 347–388.

[85] Donghee Yvette Wohn, Guo Freeman, and Caitlin McLaughlin. 2018. Explaining viewers' emotional, instrumental, and financial support provision for live streamers. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. ACM, 474.

[86] Julia T Wood. 1994. Gendered media: The influence of media on views of gender. *Gendered lives: Communication, gender, and culture* 9 (1994), 231–244.

[87] WriterDuet. 2020. *General Statistics, Gender Overview and Lines by Type.* https://help.writerduet.com/article/89-general-statistics-gender-overview-and-lines-by-type

[88] Lei Wu, Lingli Guo, Hao Fang, and Lijun Mou. 2018. Bullet Graph Versus Gauges Graph: Evaluation Human Information Processing of Industrial Visualization Based on Eye-Tracking Methods. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 752–762.

[89] Siqi Wu, Marian-Andrei Rizoiu, and Lexing Xie. 2019. Estimating attention flow in online video networks. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 183.