Investigating the Use of Synthetic Media and Real-Time Virtual Camera Filters for Supporting Communication and Creativity

by

Joanne S.L. Leong

BASc, University of Waterloo (2014)
MSc, University of Applied Sciences Upper Austria (2016)

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning
in partial fulfillment of the requirements for the degree of

Master of Science in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2021

Author .....
Program in Media Arts and Sciences
January 22, 2021

Certified by .....
Pattie Maes
Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by .....
Tod Machover
Academic Head, Program in Media Arts and Sciences

Investigating the Use of Synthetic Media and Real-Time Virtual Camera Filters for Supporting Communication and Creativity

by

Joanne S.L. Leong

Submitted to the Program in Media Arts and Sciences
on January 22, 2021, in partial fulfillment of the
requirements for the degree of
Master of Science in Media Arts and Sciences

Abstract

Virtual representations of ourselves can influence the way we think, feel and behave. While this phenomenon has been explored heavily in the realm of Virtual Reality, little is known about the utility of synthetic media and real-time camera filters to reshape the perspective we have of ourselves and our capabilities. At the same time, the prevalence and popularity of these technologies have surged, coupled with greater usage of online communication tools. Motivated by a desire for self-improvement in an age of online digital communication, this thesis aims to investigate how synthetic media and real-time camera filters can be used to influence performance in target tasks, particularly in the realm of communication and creativity. This work encompasses the results of an extensive online survey (174 respondents) regarding the professional use of video-based online communication tools. It unveiled that there is an interest in self-improvement in this context and that the self-view feature of such tools may serve as a potential channel for helpful user input. Building upon this, a user study was conducted (28 participants) in which generative AI was used to synthesize videos of participants excelling in confident communication. It was found that exposure to this form of personalized media may alter feelings of confidence and stress and be implicitly helpful to some by serving as personalized role-models and guides. Following this, a second user study was conducted to explore real-time self-image manipulation via real-time camera filters (21 participants) in video calls. It was observed that applying such filters may trigger various responses in people with regards to mood, embodiment, and creativity. This thesis extends research into the topic of self-image manipulation and opens up a novel perspective that such technologies may help to serve as accessible and scalable approaches to manipulating one's self-view to nurture personal growth.

Thesis Supervisor: Pattie Maes
Title: Professor of Media Arts and Sciences

Investigating the Use of Synthetic Media and Real-Time Virtual Camera Filters for Supporting Communication and Creativity

by

Joanne S.L. Leong

This thesis has been reviewed and approved by the following committee members

Professor Pattie Maes .....

Professor of Media Arts and Sciences

MIT Media Lab

Dr. Janet Baker .....

Research Affiliate

MIT Media Lab

Professor Ehsan Hoque .....

Associate Professor of Computer Science

University of Rochester

Acknowledgements

I firstly would like to thank my advisor, Professor Pattie Maes, for her kind support and guidance over the course of this thesis and throughout this exceptional period of time. Thank you for our regular research discussions and for infusing your ideas, energy, and enthusiasm into my work. I would also like to express gratitude towards my thesis readers, Dr. Janet Baker and Professor Ehsan Hoque who were kind, understanding, and helpful amidst the pandemic. Thank you for your input and guidance via emails and Zoom calls, which helped to shape the direction of this work and elevate it.

I'd like to thank my collaborators on these projects, Pat Pataranutaporn, Yaoli Mao, and Florian Perteneder. I am so grateful to have had the chance over the last year to work with you and learn from you. I truly valued all of the discussions we had, spanning everything from the high-level concepts right down to the gritty details of study design, implementation, and analysis. Your creativity and research skills inspire me to keep growing, and I look forward to more collaborations with you.

Warm thanks to the many people I've gotten to know during my time here at the MIT Media Lab. Thank you to everyone in the Fluid Interfaces group, including Caitlin Morris and Camilo Rojas for our summer research chats, Adam Horowitz for his insights on psychology and creativity, and Naoki Ohsugi and Andrew Wong for their help in piloting my studies. Many thanks to Roy Shilkrot and Pat with whom I had the privilege to teach the MAS.S60 Experiments in Deepfakes class. Sincere thanks to Hidefusa Okabe for shining a light on statistics. Thank you to the members of the Langer lab, especially Professor Ellen Langer for inviting us to her research meetings and Jay Olson for his helpful critique of the creativity study. Many

thanks to the undergraduate students who have worked with me, Jose Martinez and Raul Alcantara. Thank you also to my dear friends in TMG. Thank you to Professor Hiroshi Ishii for inviting me to the Media Lab and imparting the importance of having a vision. Thank you to Ken Nakagaki for the many fun research chats and moments of excitement in getting our swarm robots to dance. I'm sure you'll be an amazing professor and advisor! Thank you to my dear friends and brilliant office mates Hila Mor and João Wilbert. I couldn't have had better people to "enjoy the climb" with. Thank you to Dan Levine who was always kind and helpful, from day one. I am so happy your dreams to fly are coming true! Thank you to Taka and Yun for showing us the ropes, and Jifei, Koichi, Jack, Liang, Alice, Deema, Paula, and Jonathan for the kind support and shared laughter throughout. Also, a sincere and special thanks to Monica Orta and Kerianne Panos for being lifelines when I really needed it.

Thank you to all my friends from times before, who have stuck with me on this roller coaster and have always given me reasons to laugh no matter what comes. To lieber Flo, thank you for always being there by my side and pulling through for me in every way. I am so grateful to have you in my life, and truly, none of this would have been possible without you. Thank you for being my rock who makes the perfect everything buddy. To my family, I feel so grateful and blessed every day to have you. Thank you for always having my back. I wouldn't be who I am today and where I am today without your support and love. Everything until this point would certainly not be possible without you. A special thanks to my siblings, parents, and especially grandparents. You've made this all possible. Love you always.

Contents

Acknowledgements **7**

1 Introduction **13**

1.1 Initial Remarks 13

1.2 Motivation 14

1.3 Core Concept 16

1.4 Definitions 18

1.4.1 Synthetic Media 18

1.4.2 Real-Time Virtual Camera Filters 18

1.5 Thesis Outline 19

2 Related Work **21**

2.1 Synthetic Media & Virtual Camera Filters 21

2.2 Triggering Psychological & Behavioural Change 23

2.2.1 Plasticity of the Self: Embodiment, Enfacement & the Proteus Effect 23

2.2.2 Imitation & Mimicry 24

2.3 Confidence in Communication 25

2.3.1 Measuring & Analyzing Confidence in Communication 25

2.3.2 Interfaces to Improve Communication Skills 26

2.4 Creativity 27

2.4.1 Measuring & Analyzing Creativity 27

2.4.2 Technology-Mediated Creativity 28

3 Survey on Video-Based Online Communication **29**

3.1 Motivation 29

3.2 Method 30

3.3 Results 31

3.3.1 Demographics 31

3.3.2 Tools & Features 32

3.3.3 Types and Frequency of Calls 33

3.3.4 Pros & Cons of Video Calls 34

3.3.5 Speaking with Confidence 35

3.3.6 Improving Communication Skills 36

3.4 Implications 37

4 User Study 1: Exploring Synthetic Media for Confident Speech **39**

4.1 Motivation 39

4.2 Concept 41

4.3 User Study 42

4.3.1 Conditions 42

4.3.2 Experimental Set-Up & Apparatus 43

4.3.3 Procedure 44

4.4 Analysis & Results 46

4.4.1 Demographics 46

4.4.2 Attitudes & Impressions 46

4.4.3 Subjective Effects 49

4.5 Discussion & Limitations 51

4.6 Summarizing Remarks 55

5 User Study 2: Exploring Real-Time Camera Filters for Creativity **57**

5.1 Motivation 57

5.2 Concept 59
5.3 User Study 60
5.3.1 Conditions 60
5.3.2 Experimental Set-Up & Apparatus 61
5.3.3 Standardized Tests & Response Variables 62
5.3.4 Procedure 65
5.4 Analysis & Results 66
5.4.1 Demographics 66
5.4.2 Personality 67
5.4.3 Filter Impressions & Beliefs 67
5.4.4 Embodiment 69
5.4.5 Mood 72
5.4.6 Creativity 73
5.4.7 Implicit Associations 74
5.5 Discussion & Limitations 75
5.6 Summarizing Remarks 78

6 Conclusion & Future Work **81**

References **85**

A Online Communication Survey **97**

B Communication Study Pre-Survey **107**

C Communication Study Task Instructions **111**

D Creativity Study Pre-Survey **119**

E Creativity Study Task Instructions **127**

F Creativity Study Trial Survey **131**

G Creativity Study Exit Survey **137**

Chapter 1

Introduction

1.1 Initial Remarks

The use of video-based online communication has grown over the last decade and has surged due to the need for remote work during the COVID-19 pandemic [107]. The rapid and widespread adoption of this technology has brought to light a new challenge, namely, "Zoom fatigue" – the tiredness and anxiety that stems from the heavy use of video-conferencing systems [107]. Nevertheless, as the pandemic continues to grip many countries and many large technology corporations have announced that they will adopt remote work as a central aspect of their operations [34], it appears that this technology is a necessity that will likely become a mainstay in the post-pandemic landscape of work and productivity tools.

Human-Computer Interaction (HCI) strives to make interactions with computer systems as efficient and easy-to-use as possible so that people may realize their goals [18]. It is a highly multi-disciplinary domain that has evolved to encompass a wide variety of missions and themes, including context-aware and wearable computing, perceptual augmented and virtual reality interfaces, digital design and fabrication, data visualization and exploration, collaboration and learning and more [11]. As

an HCI researcher, one of my main interests has been to design interactive systems to help people learn and develop their abilities in order to become the best version of themselves. In the context of this period of time from 2020 to 2021, I saw and experienced first-hand the sudden shift and growing pains that came with a newfound mass reliance on online communication tools. I contemplated in what ways we may change our experiences with these tools to best leverage them for our personal benefit.

One curiosity that emerged from using these tools is the ability to see ourselves and alter our self-image. Video-conferencing interfaces not only present video streams of one's conversation partners, but also offer a literal window to see oneself. The ability to view ourselves in our interactions with others is a phenomenon that has never before been available in traditional face-to-face exchanges. This sparked my interest in the implications of a digital self-view. Compelling aspects of this are its wide degree of accessibility, as well as the ability to digitally manipulate them. In this thesis, I explore the potential opportunity space that is created through technology-enabled self-representations, and investigate how new advancements in synthetic media and the development of real-time camera filters for online communication may offer a new lens through which we see ourselves. This in turn can be used to positively influence how we think, feel and behave in our everyday endeavors.

1.2 Motivation

In a famous experiment in 1979, known as the "counterclockwise experiment" [64], an experimental group of eight elderly men experienced a five-day-long retreat of living '20 years in the past' [78]. The researchers conducting the experiment wanted to know if giving people the illusion of living in the past would not only be able to make people feel younger, but actually reverse the effects of aging. In order to establish this illusion, artifacts from the era such as magazines were embedded into their environment, and television and radio programs were made to play programs

from 20 years prior. Participants were prompted to have twice-daily discussions about events from the time in present tense, and to refrain from discussing anything that occurred beyond that time. The researchers reported measuring physiological improvements (i.e. improved hearing, vision, grip strength, memory, and more) as an effect of psychologically 'turning back the clock.' While a control group that was told to reminisce about the past and the experimental group both improved, the improvements observed in the experimental group exceeded those of the control group. This elaborate physical setup was found to change the well-being and physical capabilities of those who lived inside it.

Technological developments in virtual reality (VR) have since made it possible to digitally create immersive experiences and illusions, which has enabled researchers to push the boundaries on psychological experimentation and research. In VR, people can be given the impression of being in different physical [20, 68] or social environments [38, 20]. Interestingly, it also enables people to have the illusory experience of embodying someone else. For instance, a person can be made to embody someone who is older [45] or younger [98, 44], of a different physique [60], of the opposite gender [80, 74], or of a different race [57]. People can also embody an entirely different character, such as an inventor [40], or a famous person such as Einstein [8]. In line with the counter-clockwise experiment, such research has demonstrated that carefully designed immersive experiences can have the profound effects of changing peoples' attitudes and behaviours [33, 12, 45] and enhancing their cognitive capabilities [8, 40].

These works present evidence that generating perceptual illusions can have wide-reaching effects on our perception, behaviour and attitudes for many positive applications. While growing numbers of findings and developments in VR point to its positive potential in gaming, education and even workplace productivity [23, 92], VR has yet to become mainstream. In this thesis, I am therefore motivated to examine and harness the potential of synthetic media and real-time camera filters to change the perspectives we have of ourselves for the purposes of self-improvement.

1.3 Core Concept

Perceptual illusions have the power to change our abilities and behaviours. Prior research in psychology and virtual reality (VR) has shown that psychological and even physiological changes can be procured through shifting people's perception of their physical environment, social environment, or even their own bodies as shown in Figure 1-1 (a, b and c). In particular, VR has demonstrated that virtually embodying another character can boost people's cognitive abilities. For instance, embodying Einstein increased the cognitive task performance of some men [8], whereas engineering students became more creative when embodying inventor characters [40]. These works motivated us to investigate if this concept of digitally altering one's self-image could also be applied in the context of remote interactions over video-conferencing. With present-day tools, people are shown live camera streams of their conversation partners as well as themselves. This **self-view** opens a unique opportunity for our digital self-image to be digitally synthesized or even manipulated on-the-fly.

Based on this idea, this thesis aims to explore the utility of **digital synthesis and manipulation of one's self-image** in the context of online and remote video-based interactions, extending the existing research in two directions. Firstly, we explore the impacts of digitally changing one's **visual appearance** using **real-time**

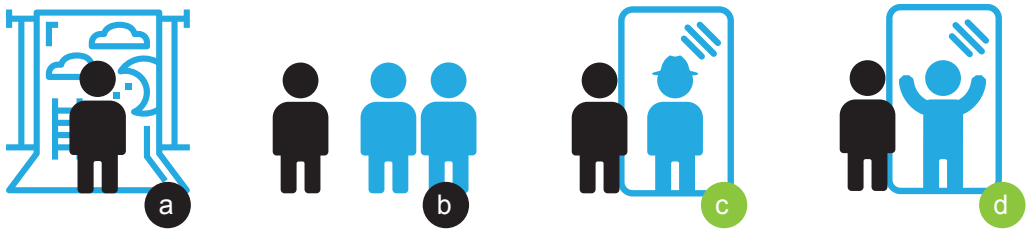


Figure 1-1: Simulating different physical environments (a), and social environments (b) has been demonstrated to have effects on our cognition, behaviours and physicality. In this thesis, we explore novel methods for and the effects of manipulating visual self-image (c) and the perception of one's own abilities (d), marked in green.

virtual camera filters (refer to Figure 1-1, c). While this has been done extensively using VR technology, to our knowledge, this possibility has not been explored before using virtual camera filter techniques. Secondly, we propose the novel concept of manipulating the perception of one's own **abilities** using state-of-the-art **machine learning-based synthetic media techniques** (refer to Figure 1-1, d). We argue that this is a new dimension along which to manipulate one's self-perception.

In exploring the merits of altering digital self-image during video-conferencing, we focus on applying it to the domains of *confident communication* and *creativity*, which are both valuable and important traits in the workplace. Augmenting creative and communicative abilities has been the core focus of many research efforts within HCI for which many approaches have been developed. Categorizing these efforts, we can distinguish between implicit and explicit cues used for guidance. As we aim to synthesize or alter one's digital self-image, we take the approach to offer implicit guidance. Doing this, we rely on psychological phenomena such as mimicry and the embodiment illusion to transfer skill. In line with this, we aim to alter one's attitude and mindset entering these activities, rather than address specific low-level behaviours, such as pitch or speech rate in the case of confident communication.

Summarizing, in this thesis, we aim to utilize the self-image within video conferencing solutions to help people improve their skills in communication and creativity by providing feedback to shape a person's attitude and mindset. These efforts are informed by an extensive online survey investigating the use of state-of-the-art video communication tools in professional settings presented in Chapter 3. Based on this, in Chapter 4, we explore whether people's beliefs about their own abilities can be changed through exposure to personalized synthesized videos that show them excelling in this skill. We apply this concept in the context of confident speech and investigate its impact on people's subjective perception of their confidence and performance while speaking. In Chapter 5, we explore whether changing a person's visual appearance to themselves in real-time can influence their self-perception and consequently their

creative cognition. By visually embodying different characters (i.e. an inventor or child), we explore whether people's performance changes in accordance with their change in appearance. These efforts provide a first exploration of the opportunities and limitations that come with leveraging one's self-image in video conferencing tools for enhancing abilities. Moreover, we hope that this research can highlight positive capabilities of real-time video filters and synthetic media techniques, which are often only considered gimmicks for entertainment purposes or even harmful.

1.4 Definitions

In this thesis, two user studies were conducted based on the use of *Synthetic Media* techniques and *Real-Time Virtual Camera Filters*. These terms are defined below.

1.4.1 Synthetic Media

Synthetic media refers to any data or media (e.g. visual and audio artifacts) that have been synthesized using algorithmic means, especially through the use of artificial intelligence (AI) techniques [97]. Alternative terms for this include *AI-generated media, generative media, and personalized media* [97]. *Deepfakes* are a subset of synthetic media in which people appear to say or do something that they have not.

1.4.2 Real-Time Virtual Camera Filters

Real-Time Camera Filters refer to virtual filters that can be applied to one's camera stream in order to change its visual appearance. They have become increasingly pervasive in mobile social media applications (e.g. Instagram [49], Snapchat [91]), allowing people to take photos or videos with artificial visual effects instantly applied to their camera streams. Applications such as Snap Camera [90] and Zoom [112] have

also provided the capability of applying these visual effects in real-time for online-based communication [113] on laptop and desktop computers.

1.5 Thesis Outline

This thesis is comprised of 6 chapters. Following this introduction, the related work is presented in Chapter 2. After this, we describe the results of an online survey in Chapter 3, which serves as the motivation for two user studies presented in Chapter 4 and 5. Contributions and insights from these explorations as well as ideas for future work are summarized in Chapter 6 to conclude this thesis.

Chapter 2: Related Work forms a foundation by outlining technical aspects and psychological theories. It also discusses relevant works in the domains of communication and technology-mediated creativity.

Chapter 3: Survey on Video-Based Online Communication presents the insights of an extensive online survey on the current use of video-based communication tools. In particular, it investigates requirements with regards to confident speech and highlights opportunities for technological interventions based on self-view.

Chapter 4: User Study 1: Exploring Synthetic Media for Confident Speech investigates the utility of short video snippets showing the users themselves speaking confidently prior to answering questions in a user study run with 28 participants.

Chapter 5: User Study 2: Exploring Real-Time Camera Filters for Creativity presents an effort to better understand the impact of real-time video filters on embodiment, mood, and creative tasks. In a study with 21 participants, 2 filters and a baseline condition without a filter were applied to explore alteration of self-image during video calls.

Chapter 6: Conclusion & Future Work summarizes the insights and contributions of this thesis and highlights a number of avenues for future work.

Chapter 2

Related Work

In this thesis, we explore how changes to self-perception induced using synthetic media and real-time camera filters can impact performance in communication and creativity. We therefore provide an overview of the respective technologies, outline psychological phenomena that may serve as mechanisms for how exposure to these technologies may have impact, and lastly outline works pertaining to improving people's capabilities in the given application domains.

2.1 Synthetic Media & Virtual Camera Filters

Synthetically produced virtual faces have been used to improve conversation quality [61], serve as virtual audience members for practicing public speaking [89], and train social skills [10].

Advancements in machine learning have enabled a new generation of algorithmically generated hyper-realistic synthetic media, spanning images, text, audio, and video. Generated artifacts featuring the human face, body, and/or voice that are realistic enough to be perceived by viewers as authentic have come to be known as *deepfakes* [70]. Many approaches exist to synthesize visual deepfakes [75, 88], such as

using autoencoder-decoder pairings, generative adversarial networks (GANs), as well as neural networks in combination with compositing techniques [96]. These visual methods yield surprisingly realistic results even with just minimal input, for instance even with a single image of a target person [88]. Methods also exist for voice synthesis and cloning [52, 104, 4]. However, many hurdles remain to create convincing voice samples. Most models require extensive training with multiple samples and high-quality audio recordings to attain relatively higher levels of audio quality. However, on average, the output from these algorithms are still easily discernible from natural voices, particularly if they rely on minimal audio input (e.g. a 10-second audio clip from a target speaker) as can be done with [52].

"Deepfakes" have a widespread reputation for malicious use. This technology has given rise to a host of ethical concerns, primarily with its potential use for defamation and disinformation [79]. For instance, celebrities and politicians can be portrayed making statements or performing actions that they did not engage in in real life. Consequently, there are growing efforts to correctly detect synthetic media [2]. On the flip side, synthetic media also has the potential to be used for positive applications. For example, they can be used to power engaging interactive experiences [24], overcome challenges resulting from injury or illness [26], rally support for important causes [25], or provide new and critical perspectives on historical events [5]. In Chapter 4 of this thesis, we aim to change a person's self-perceived abilities by transferring the skill of another speaker into the person's likeness using synthetic media technology.

Real-time virtual camera filters can manipulate a live camera stream to change the appearance of a person. They are becoming widely accessible across a wide range of social media and online communication platforms [90, 112]. Many filters rely on computer vision techniques to track facial features and manipulate the look of a person by overlaying visual effects. As such, filters have traditionally produced effects that have a mask-like quality. However, more recently, these filters in combination with machine learning algorithms can produce highly realistic alterations to one's digital

appearance in real-time. Consequently, these effects (e.g. making someone look like they are younger or older, wearing make-up, have different hair) are becoming harder for people to discern as being fake. In this way, this technology stream is converging with synthetic media, as it can be used to make people appear so different as to be unrecognizable or look like somebody else altogether (yet look highly realistic).

2.2 Triggering Psychological & Behavioural Change

Many external influences can shape how we feel, think, and behave. Researchers have sought to understand how this occurs and identify potential mechanisms for change. We delve into topics that relate to how synthetic media and real-time filters may impact users in the context of having a self-view in online communication scenarios.

2.2.1 Plasticity of the Self: Embodiment, Enfacement & the Proteus Effect

Our sense of body ownership is surprisingly flexible and can be altered given congruent sensory inputs as stimulation. The famous "rubber-hand illusion" (RHI) [14] showed that the brain can substitute a rubber-hand as being one's own given (1) the tactile sensation of having one's real hand brushed and (2) the visual input of the rubber hand being stroked at the same time as one's real hand (hidden from view). Multisensory areas in the brain integrate the input we receive through our senses to give salience to events that happen to and around us [94, 32].

In virtual reality, multisensory integration is used to create bodily illusions on avatars; both visuo-tactile and visuo-motor integration techniques are used to trigger an embodiment illusion – the belief that an avatar body is one's own. This illusion can be applied to parts of a body, the whole body, and even the face [103, 69]. The experience of embodiment is complex and encompasses a sense of *body ownership* as

well as *agency, co-location, and external appearance* [36]. As these experiences are personal and vary across participants, researchers have formulated a questionnaire to establish a standardized approach to measure the extent of the embodiment illusion [36]. Virtual embodiment is the foundation of many social VR experiences and has been shown to alter performance and behavior in a variety of ways [8, 57, 81, 27]. In fact, the Proteus Effect [109] is a well-established phenomenon wherein our behaviors conform to our digital self-representations. This effect has implications beyond VR and has been shown to influence people's behavior in online gaming scenarios, and even in subsequent face-to-face interactions [110]. In Chapter 5 of this thesis, we explore how real-time virtual camera filters may induce the embodiment and enfacement illusions to boost people's performance in creative tasks.

2.2.2 Imitation & Mimicry

Imitation and mimicry are considered "pervasive and automatic in humans" [48]. They are believed to shape our social and learning behaviors [108] and are driven by mirror neurons in the brain [48]. Emulating the speech or physical behavior (e.g. gestures, facial expressions) perceived in another person is common [30], while experiments in priming have shown that imitation can also be more complex and subtle [30]. With priming, both conscious and unconscious exposure to one stimulus can influence a person's response to a subsequent stimulus [7, 51]. Its effects have been investigated in many contexts from sports [6] to creativity [66]. Furthermore, priming can impact both linguistic and paralinguistic features of prosody [102] such as speech rate [55], intonational phrase structure [101], and even sentence structure [111]. In the user study described in Chapter 4 of this thesis, we make a novel exploration into whether aspects of speech and confidence can be transferred through imitation and mimicry of a synthesized video of oneself speaking confidently.

2.3 Confidence in Communication

Confidence is a critical element in successful communication that is felt both by the speaker and the listener. In the field of psychology, defining confidence in the context of speech remains a challenge. Kimble and Seidel [59] describe it as the "feeling of knowing or unknowing". Jiang and Pell [54] define it as the "demonstrated commitment to the content of their speech," which is transient in nature. Conversely, Tenney et al. [100] assert that it is a static mental state that pertains to one's social status or personality. Confidence in communication has also been investigated with regards to semantic content, low-level acoustic, and visual features [59, 54, 72]. In Chapter 4, we examine confidence through the lens of users who are asked to speak impromptu in an interview context, using self-reports from speakers.

2.3.1 Measuring & Analyzing Confidence in Communication

Confidence in speech can be subjectively felt by a speaker as well as detected by a listener. Humans have a remarkable ability to decode how much a speaker feels confident about what they are saying; a study by Jiang et al. suggests that an initial judgement about a speaker's confidence can be made 200 milliseconds after the onset of speech [53]. However, how does one detect or measure confident speech? Researchers have sought to pinpoint cues that may indicate confidence to listeners and observers, concluding that confidence can be encoded both by what you say, and how you say it. Kimble et al. [59] and later Jiang and Pell [54] have analyzed acoustic features and their relation to confidence. In their study based on single sentence statements, they found that distinct patterns of pitch, intensity, and temporal features corresponded with different confidence levels. In general, confident expressions had the highest fundamental frequency (f0) range, mean amplitude, and amplitude range, while unconfident expressions were characterized by having the highest mean f0, slowest speaking rate, and more pauses. Lexical cues (e.g. "I'm certain" vs. "I'm

pretty sure") were found to amplify the intended level of confidence. Mori and Pell [72] studied visual markers that change the perception of confidence and found that speakers with high confidence in their speech content were more likely to maintain eye contact, a still posture, and a serious facial expression, whereas an upwards gaze and thinking facial expression corresponded more with low levels of confidence. Observers accurately perceived a speaker's confidence level relative to the speaker's subjective feeling of confidence based on the speaker's eyes, facial expressions and head movements as cues.

2.3.2 Interfaces to Improve Communication Skills

In HCI, several works have been developed to help people improve their communication skills by giving both post- and real-time feedback to users about their behaviours. Wearables such as Logue [28] and Rhema [99] offered real-time feedback to speakers to help them modulate their speech rate, gestures, speaking volume, and more. MACH [46] was an interview practice system comprising a virtual agent and analytical system that summarized users' speaking habits. Researchers also explored the use of a smart speaker conversational agent to manage a user's public speaking anxiety [106]. With regards to group dynamics, Coco [87] would track and highlight how people respond to each other in online video conferencing sessions, and SociaBowl [65] would serve as a tangible platform to help mediate group conversations. Platforms have also been developed as assistive technology. SayWAT [16] and vrSocial [15] assist people with autism in providing real-time cues on physical proximity or volume in social situations. A range of commercial apps have also been developed to improve speaking and presentation skills (e.g. ORAI [77], Presentr [84], Speeko [93] and PowerPoint Presenter Coach [82]) that track aspects ranging from speech rate to the use of filler-words. In contrast, we explore in Chapter 4 of the thesis how one's performance and feelings can be influenced by synthetic media in an impromptu speaking scenario.

2.4 Creativity

Creativity has been defined as the ability to discover novel and useful solutions to a given problem [3, 86]. It is often mapped to divergent thinking abilities [42, 50]. Many tests and techniques have been created to measure and assess one's creative thinking abilities [42, 58, 83, 50]. In Chapter 5 of this thesis, we use two standard creativity tests and automated assessment techniques to determine the potential impact real-time camera filters may have on creativity.

2.4.1 Measuring & Analyzing Creativity

One aspect of creative thinking is divergent thinking. The Alternate Uses Task (AUT) is a well-known method for assessing one's divergent thinking ability proposed by Guilford et al. [42]. The original variables [42] for evaluation include *Originality, Fluency, Flexibility and Elaboration* [42]. While the results from the AUT are typically analyzed manually by human raters, [105], automated methods for assessing responses are being developed [9]. One method to automate creativity assessment is by calculating a semantic distance between pairs of text [83] using latent semantic analysis (LSA). Larger distances between two texts, such as a prompt and a response, indicate that they are conceptually farther away and that the response is more creative. SemDis [9] is a freely accessible online web tool that uses this technique as a basis to assess creativity based on responses from an AUT. In this thesis, we rely on automated analysis of creativity for the AUT and also a Verb Generation Task (VGT). The VGT is another creativity assessment designed by researchers that involves asking participants to state a verb when prompted with a noun. Semantic distances between the noun-verb pairs have been shown to be an indicator of greater creativity [83]. In Chapter 5 of this thesis, we leverage the AUT and VGT to capture and assess the influence different real-time virtual camera filter conditions may have on a person's cognitive performance in creativity.

2.4.2 Technology-Mediated Creativity

In human-computer interaction (HCI), the prospect of augmenting one's creative abilities has spurred many unique research efforts [35, 47]. Nakazato et al. [73] investigated how changing the facial appearance of pairs of people, either through expression or similarity between them during video calls, impacts the outcome of collaborative brainstorming. In online crowdsourced environments, computational priming via the assumption of different roles coupled with affective stimulation via images was studied as a means to boost creativity, and was found to be potentially helpful when one runs out of ideas [76]. *Avatar-mediated brainstorming tools* [41, 17, 67] are an emerging class of creativity support tools. In a study by Guegan et al. [40] subjects who experienced embodying "inventor" avatars in VR demonstrated greater fluency and originality of ideas in a collaborative brainstorming task than those in the neutral avatar and standard face-to-face control conditions. Avatars and creative idea generation have also been studied by Marinussen and Rooij [67], who proposed that self-similar avatars offer a creative boost by increasing feelings of self-identification and positive affect. In contrast to these works, we perform a novel investigation into the potential for real-time camera filters to alter self-image, by allowing subjects to see themselves as an inventor-like figure or as a child. The details of this exploration are presented in Chapter 5 of this thesis.

Chapter 3

Survey on Video-Based Online Communication

In order to understand the potential design space for interventions within the context of video-based online communication, we conducted an extensive online survey on the topic. In this chapter, we elaborate on the motivation, methods and findings of this survey. The outcomes formed the foundation for the subsequent explorations conducted as part of this thesis in the domains of communication and creativity.

3.1 Motivation

The onset of the COVID-19 pandemic quickly unearthed the need to rely heavily on online communication tools, particularly for work purposes. This opens a unique opportunity space to potentially provide helpful interventions to aid in communication and collaboration due to the digital nature of this mode of communication. In traditional face-to-face interactions, it is challenging to deploy helpful technological interventions; this can be accomplished to some degree using wearable technologies, but even platforms with minimal designs such as Google Glass face challenges with

social acceptability – particularly in the context of in-person interactions [63, 62].

In order to design tools for self-improvement that are helpful, timely and relevant in the current situation, we were firstly interested in understanding people's current experiences with using online video-based communication tools for work. For instance, we wanted to know which platforms and devices people use, what types of calls they engage in, where people's gaze is drawn to during conversations, what are the pain points people have with this mode of communication, what are the cornerstones for effective communication, and what kinds of interventions or feedback if any people would be interested to have. These questions motivated us to conduct an online survey. As we were primarily focused on learning and skill-development, we focused our study towards video communication in professional settings since high performance is critical in these contexts.

3.2 Method

The study was approved by the Committee on the Use of Humans as Experimental Subjects (COUHES), that serves as the institutional review board (IRB) for the Massachusetts Institute of Technology (MIT). It was created and disseminated using Google Forms. People were invited to complete the survey on a voluntary basis. The survey was distributed via emails, public communication platforms and invites over various direct messaging platforms. Each participant gave consent for their information to be collected for research purposes. The survey comprised several questions in various formats including multiple choice questions, rating scale questions and open questions. Questions were categorized into different sections in accordance to their respective themes: *platforms and features, speaking with confidence, types of calls, and demographics.*

3.3 Results

Here we describe the results of the survey on the experience of **using video-based online communication tools for work**. This includes insights into the types of tools and features people use, the calls they engage in, and potential opportunity spaces people identified for receiving feedback to improve their communication skills.

3.3.1 Demographics

A total of 203 people responded to the survey. As we were specifically interested in people's use of video-based communication platforms for *professional activities*, 29 respondents were screened out. In the end, there were 174 people who responded to the full survey, of whom 96 were male, 76 were female and 2 were undisclosed. Their ages ranged from 17 - 80, with the average age of all participants being 43.5, average age for males being 47.1, and average age for females being 39.1. The age distribution is shown in Figure 3-1. Respondents included people with various occupations working in different sectors including finance, health, science and technology, education, and more. Examples of occupations held by the respondents include accountants, bankers, investment advisors, physicians, therapists, psychologists, engineers, software designers and developers, managers, teachers, professors and students.

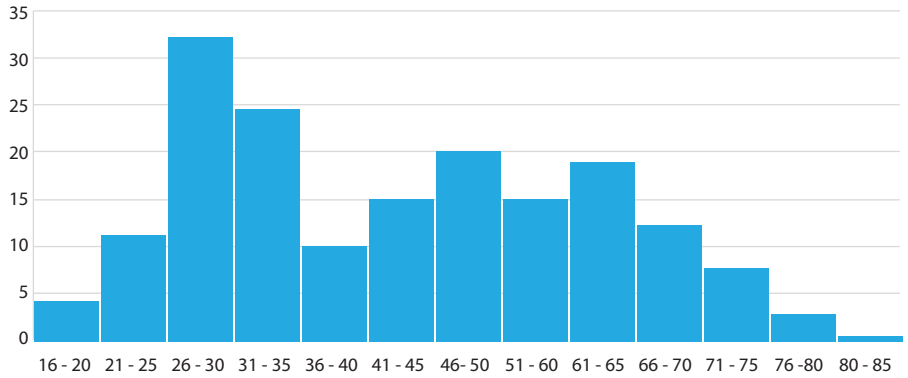


Figure 3-1: The distribution of ages of respondents in the online survey.

3.3.2 Tools & Features

People were asked to indicate which devices they use for professional video-calls. The options were not exclusive. People predominantly used their laptops (91.9% of all respondents) to hold video-calls, followed by mobile devices and tablets (33.7% and 12.2% respectively). They reported using a wide variety of video-calling platforms for work. The top five most commonly used platforms amongst our respondents were Zoom, Skype, Microsoft Teams, Google Hangouts, and WhatsApp (see Figure 3-2).

Participants were asked to report their habits around watching the video streams of others as well as watching their own self-view. For the self-view, 7% *never look at it*, 14% only look at it *in the beginning*, the majority of respondents look at it *once in a while* (46%), *quite frequently* (28%), and *almost constantly* (5%). Therefore, the majority of people (78%) are in the habit of looking at their self-view during video calls at least every-so-often.

Reasons for looking at one's self-view are summarized in Figure 3-3 (right). More than half of the respondents reported using the self-view to check their physical

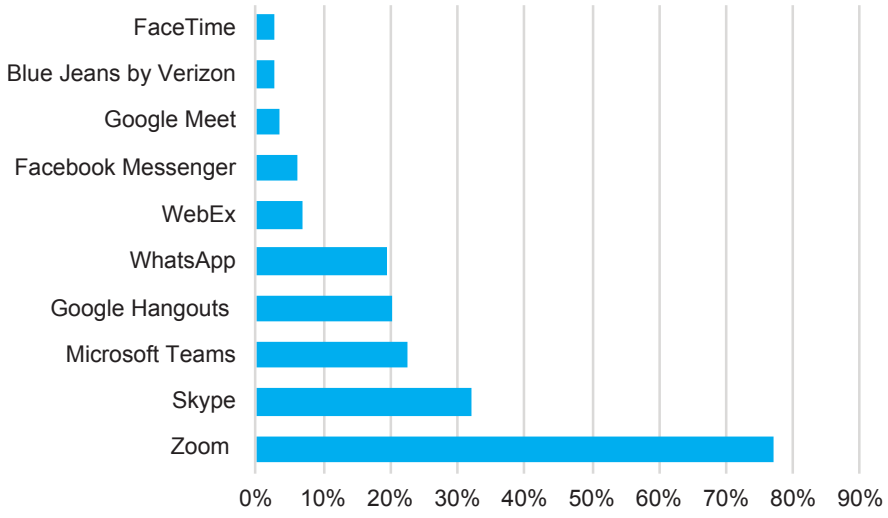


Figure 3-2: Percentage of respondents using different communication platforms.

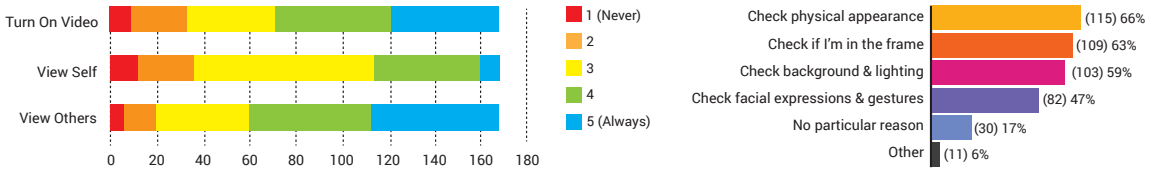


Figure 3-3: Results from online communication survey. Left: summary of use of different features (1 = Never, 5 = Always). Right: reasons respondents have for checking their own self-view as a percentage of all respondents. Note: the responses of 7 participants were dropped for this analysis due to contradicting answers (reported *Turn On Video* = 1 but *Viewing Self* > 1)

appearance, background, lighting, and whether they're in the frame. Just under half reported using it to check their facial expressions and gestures. The survey highlighted that in the context of video-based online communication, there is the potential for one's self-view to serve as a channel for helpful input (from system to user). Given that people are able to see an image of themselves during video calls, real-time filter effects in this context may be able to play a constructive role. The majority of our participants sometimes or often use screen-sharing. When screen-sharing is being used, most people never or mostly never look at their self-view.

3.3.3 Types and Frequency of Calls

People were asked how often they use a video-based communication platform. 3.4% said *continuously*, 21.8% reported *multiple times daily*, 17.8% use it *every day*, 38.5% reported using it *a few times each week*, 18.4 % use it *less than once a week*. As such, the vast majority of respondents (approx. 78.8%) reported using video-based communication platforms for work *at least a few times each week*. Broken down by meeting-size, it was most common for people to have 1-on-1 and small group calls (<5 people) a few times a week, medium-sized group calls (6-10 people) once a week, and large to very large group calls (11-25, >25 people) less than once a week.

The top five most common types of calls people engaged in were *information sharing* (81%), *status updates* (79.3 %), *planning/decision making* (72.4 %), *problem*

solving/brainstorming (65.5 %), and *team building* (47.1%). For their meetings, respondents were asked if they must prepare for speaking or speak impromptu. For impromptu speaking, answers were *never* (5.7%), *less than once a week* (23%), *a few times a week* (39.7%), every day (15.5%) and *multiple times a day* (16.1%). Thus, the majority of respondents engage in impromptu speech at least to some degree, with most people speaking impromptu *at least a few times a week or more* (71.3 %).

Respondents rarely engaged in rehearsed types of speech, with almost half (47.1%) *never* needing to rehearse. However, 82 respondents (42.4%) reported needing to rehearse *less than once a week*, 14 (8%) do so *a few times a week*, 3 (1.7%) reported rehearsing *every day*, and 1 (0.6%) person reported rehearsing *multiple times a day*.

3.3.4 Pros & Cons of Video Calls

People were asked to report reasons they may be dissatisfied with their communication abilities in the context of video-based calls. In open comments, pain-points that arose were *turn-taking* without interrupting one another, *maintaining a good conversational flow* and *managing uncomfortable silences*, *having spontaneous conversations*, and *receiving and delivering non-verbal cues*.

For non-verbal communication, respondents stated that it's harder to connect with and read others since (1) you cannot maintain eye-contact, and (2) body-language cues such as facial expressions, pointing and gesturing are less noticeable over video. To compensate for the diminishing effect, some respondents reported deliberately exaggerating how they emote. The negative consequences were that people felt it is generally more challenging to (1) understand, build and maintain engagement, attention and interest, (2) build rapport with and emotionally connect with others, and (3) maintain a natural conversational flow.

According to respondents, pros of video-calls are that it's easier to schedule and attend meetings (since physical location does not matter and no commuting is nec-

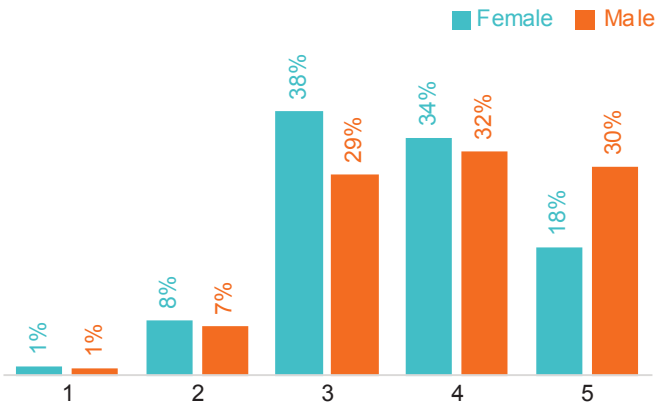


Figure 3-4: The degree to which respondents reported feeling satisfied with their communication abilities (1 = 'Not at all satisfied', 5 = 'Completely satisfied'), as a percentage of all respondents of the specified gender.

essary) as well as quickly share relevant information. Some people mentioned that video-calls help them to communicate for the reasons that they can preserve their personal space, and they find it less intimidating since physical factors are eliminated and one does not have to see everyone at once. Other aspects that people liked included the ability to 'hide', multi-task during a call, and dress more casually.

3.3.5 Speaking with Confidence

Participants were asked to report their feelings of confidence while speaking online. On a 5-point scale, the vast majority of respondents (153, 87.9 %) expressed the belief that it is either important or very important to be able to speak confidently.

Most people reported feeling somewhat or mostly satisfied with their communication abilities, suggesting that they feel there could be room for improvement. Ratings for satisfaction are pictured in Figure 3-4. We note that the distributions between genders is different, with a higher percentage of males being very satisfied with their communication skills.

People were asked about their strategies to look and sound confident. Answers included linguistic efforts such as carefully choosing one's words, structuring one's

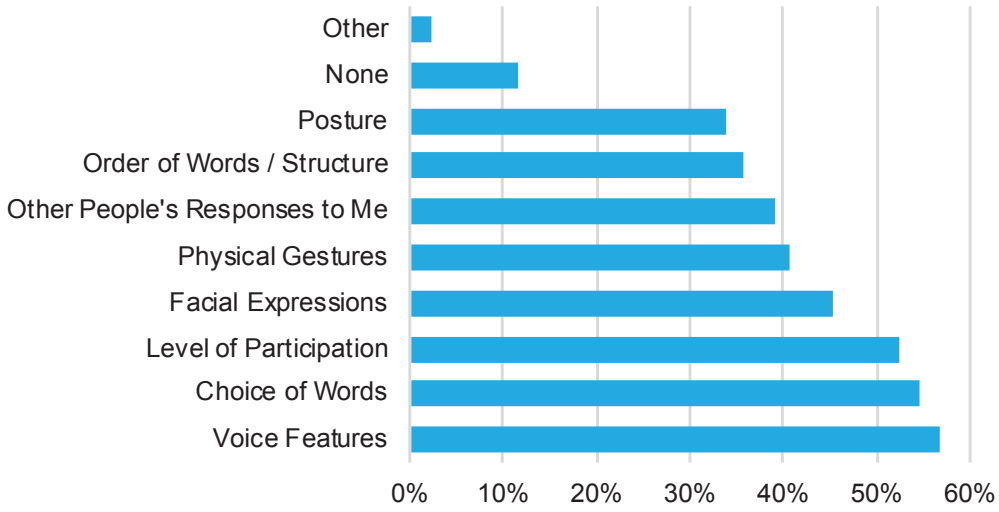


Figure 3-5: Types of information people expressed wanting to know in order to improve their communication skills in online video calls.

ideas, speaking to-the-point, and avoiding filler words or sounds. With regard to vocal delivery, people would try to project their voice, incorporate vocal variety, alter their pace, and speak with a certain tone (e.g. authoritative, intelligent, certain, assertive, purposeful or calm). Body language cues such as maintaining good eye contact and posture, smiling, and using fluid gestures without fidgeting were also said to be important to exude confidence. High-level strategies were to be present, responsive and focused on others, and well-prepared. Lastly, a professional appearance and setup (i.e. lighting, background) were also factors to come across as confident.

3.3.6 Improving Communication Skills

People were asked to report whether they would like suggestions for improving their speaking skills in video calls. 83 (47.7 %) respondents leaned towards yes, 55 (31.6%) leaned towards no, and 36 (20.7%) were unsure. They were asked what types of information would be useful to know, in order to help them speak clearly and confidently. A selection of these features are shown in Figure 3-5.

Questions were asked regarding what forms of guidance users of video-conferencing tools would be interested in. Regarding timing, 69 (39.7%) people desired advice/tips before a conversation, 65 (37.4%) people desired suggestions during a conversation, and 97 (56.3%) wished for a summary of speaking performance after a conversation.

With regard to feedback modality preferences (for suggestions only perceived by the speaker), 91 respondents (52.3%) expressed interest in *visual* hints or cues, 60 respondents (34.5%) were interested in *audio* hints or cues, followed by 45 respondents (25.9%) who were interested in *haptic* hints or cues. One respondent raised a concern that audio cues would be interesting only if it could be designed to not be distracting while in conversation with someone.

3.4 Implications

This survey established a few key takeaways that informed the design of the explorations carried out in this thesis. Firstly, the majority of people engage in video-calls using their a desktop or laptop device, and it is therefore appropriate to focus on designing helpful interventions with this device in mind. While not all people leverage the *self-view* feature, the majority of people do look at themselves at least occasionally during video calls. It is therefore an interesting channel that could be leveraged to provide input to speakers in video-conferencing situations.

The majority of people believe it is important to be able to project confidence in their video calls, and many would also be open to suggestions for improving their speaking skills in this context. Female respondents to the survey expressed being less satisfied than the males in their communication skills. This points towards potential nuances between genders regarding communication experiences. While the survey respondents expressed an interest in multiple modalities for feedback (i.e. video, audio, haptic), *visual feedback* is by far the most favoured modality. This is likely because it is less likely to distract from conversations compared to audio input (which

would use the same cognitive resources), and feedback in this form would likely be easier to interpret than information coming through haptic feedback channels. It is also an ideal modality for providing feedback since it does not require additional equipment beyond the device being used for video-conferencing.

Chapter 4

User Study 1: Exploring Synthetic Media for Confident Speech

Synthetic media open up exciting opportunities to alter people's self-perceptions. In this chapter, we present a novel idea to leverage the creation of "deepfake" videos to alter one's perception of one's own behaviours and abilities. We envision the potential for personalized artifacts and experiences to (1) boost people's confidence and (2) serve as a positive role model for them to mimic to consequently improve their performance in desired tasks. We investigate this concept in the context of communication and conduct a user study where participants watch generated videos of themselves excelling in confident communication prior to a speaking task. We explain the details and findings of this study and summarize by discussing the potential role synthetic media technology can play in assisting with self-improvement.

4.1 Motivation

The ability to speak clearly and confidently is an integral skill in almost all facets of one's life, both personal and professional. The COVID-19 pandemic has placed

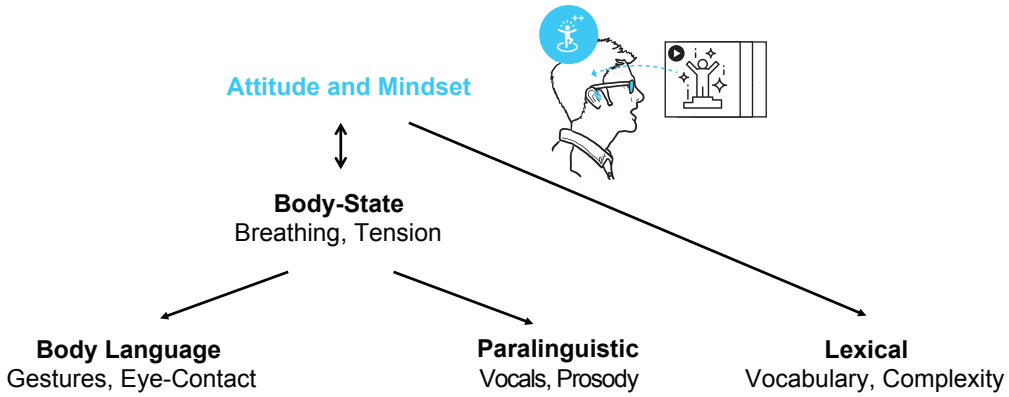


Figure 4-1: Feedback to improve communication skills can be given at different levels. Explicit low-level instructions target aspects a person can consciously control. More subconscious feedback may be able to target someone's physical state to improve their speaking quality. A system may also try to alter a person's attitude or mindset, which is the approach taken in this thesis.

greater emphasis on being able to communicate effectively via online video-based communication platforms. Responses elicited from our online survey on the use of online communication tools in Chapter 3 underline the value people place on projecting confidence and highlight a desire for tools that can help people improve their communication skills in this context.

The current landscape of tools to improve communication skills broadly fall into two categories: *practice and preparation tools*, and *real-time feedback tools*. Practice tools typically analyze one's speech across multiple features and provide a summary for people to reflect upon to improve their subsequent performance [46, 87, 77, 93]. Real-time feedback tools similarly analyze features and provide explicit pointers to be interpreted and acted-upon on-the-fly [16, 28, 43, 99, 82]. Feedback can be further subdivided into explicit cues (e.g. 'speak slower' [99]) or implicit cues (e.g. rhythmic haptic input [22]), and can tackle the problem from various levels, as pictured in Figure 4-1 to either manipulate one's behaviours [16, 28, 43, 99], one's physical state [22], or one's mental state [106]. However, as discovered in the online survey, people mostly engage in *impromptu* modes of speaking rather than practice and rehearse

beforehand. Furthermore, speaking is a cognitively demanding task. Real-time and explicit notifications in this context can add to the cognitive load and be distracting [16]. In light of this, we were motivated to find a third approach.

Inspired by the opportunities created by state-of-the-art synthetic media techniques, we set out to investigate whether short videos of confident speakers given to people before entering a conversation can be helpful by (1) serving as positive examples for people to mimic (refer to Section 2.2.2) and (2) positively influencing people's beliefs about their own abilities (as highlighted in Figure 4-1). In this way, we avoid adding to the cognitive load during a conversation, and also provide implicit guidance for self-improvement rather than bombard people with multiple explicit instructions.

4.2 Concept

In this chapter, we present the concept of synthesizing and presenting a short video snippet to a person featuring him or herself speaking confidently, prior to entering a meeting. This, for instance, could be integrated into the virtual waiting room experience of a video-conferencing system. This could serve a two-fold purpose. Firstly, it serves as a personalized positive example and role-model for a person to mimic (refer to Section 2.2.2) that is more relatable and thus easier to mimic than another person. Secondly, it may prime the person and positively influence the attitude and mindset the person carries into the meeting by changing the person's perspective of his or her own abilities.

To explore the utility of this idea, we conduct a user study, in which participants watch videos prior to answering interview questions in an impromptu speaking format. As pictured in Figure 4-2 participants are presented with (1) an abstract video, (2) videos of others speaking clearly and confidently, and (3) generated videos showing the participants themselves speaking with clarity and confidence. A generated video only comprises synthesized visuals, whereas the accompanying audio stream is taken

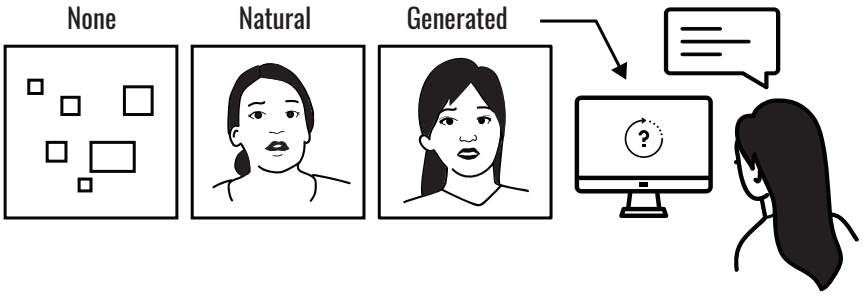


Figure 4-2: In a user study, we explore the relative impact personalized (Generated) speaking videos can have on users' subsequent impromptu speech compared to (Natural) videos of other speakers and abstract videos with no one (None).

directly from the original video source. These conditions were chosen to (1) capture a baseline, (2) assess the utility of simply seeing and hearing any confident speaker, and (3) isolate the specific impact of watching oneself speak confidently. Based on self-assessments and qualitative feedback from the participants, we observe the videos' impacts on people's feelings of confidence, stress, and performance. We summarize our findings and elaborate on interesting directions for future work.

4.3 User Study

In this section, we describe the details of the user study we conducted, including the conditions, implementation details and the study procedure.

4.3.1 Conditions

Each participant experienced three distinct video conditions in one session: **None**, **Natural**, and **Generated**. Participants were presented with two examples per condition. In the *None* condition, participants watched an abstract video that featured shape and movement but did not show anyone talking. In the *Natural* condition, the video featured another person speaking professionally. Participants would watch a video of a person of the same gender as they identified with. In the *Generated* condi-

tion, participants watched a synthesized video of themselves speaking professionally, driven by the performance of the same professional speakers. The videos comprised professional young to middle-aged adult speakers answering interview questions[1] [2] [3], or stating opinions on professional life topics [4]. Two professional speakers (one Black and one Caucasian) were sourced per gender.

4.3.2 Experimental Set-Up & Apparatus

The complete study procedure was conducted remotely online. The consent forms and pre-surveys were issued using Docusign and Google Forms. For the main study, participants were asked to complete all activities in a quiet private place using a personal desktop or laptop device with a functional webcam, microphone, and speakers (or headphones). Participants were emailed unique links to the study website, which they could access at any time they were ready. They completed the study independently without the presence of a study coordinator or live interviewer.

The website guided participants through each phase step-by-step. The website would automatically transition to the next step after playing to the end of a video or after the time ran out for answering an interview question. For the remaining activities, the participant's input was required before proceeding. All instructions and questions were given as text for participants to read. A progress bar indicated their overall progression through the study. When answering interview questions, their own camera stream was made visible on-screen, and a time progress bar was shown to

[1] Why Should We Hire You? - Sample Answer (Mid-level / Mid Career) by Big Interview:
`https://www.youtube.com/watch?v=TMxWloPWbww&ab_channel=BigInterview`

[2] Why Are You Interested In This Job - Sample Answer by Big Interview:
`https://www.youtube.com/watch?v=oA5oEr5Id5I&ab_channel=BigInterview`

[3] Describe Your Current Job Role - Sample Answer by Big Interview:
`https://www.youtube.com/watch?v=mzVXoIAsdSk&ab_channel=BigInterview`

[4] 5 things I learned from my twenties by Matt D'Avella:
`https://www.youtube.com/watch?v=ZPy8sgXuK98`

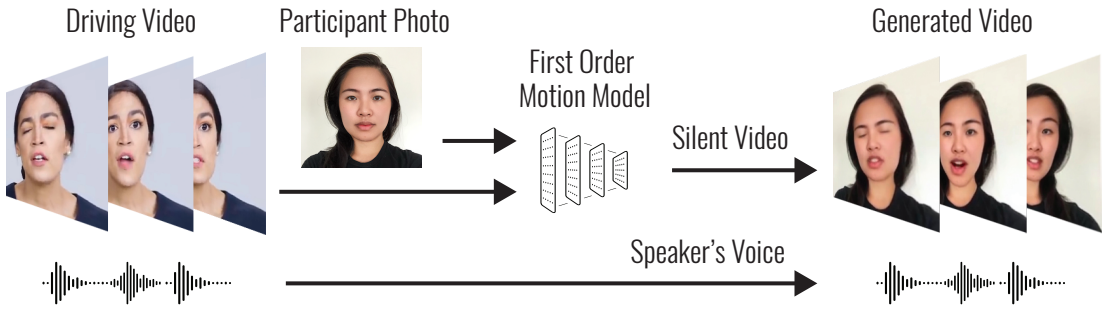


Figure 4-3: Personalized confident speaking videos are generated using the First Order Motion Model technique by Siarohin et al. [88]. The frames on the left are stills from a real recorded video of a speaker (Natural Video). The frames on the right are from a video synthesized using the algorithm (Generated Video).

indicate the time remaining. Their responses were audio and video recorded. Survey questions were built into the website, in the form of radio buttons and text boxes.

Generated videos of the subjects speaking professionally were created based on the technique developed by Siarohin et al. [88]. The pipeline to create our generated videos for each participant is pictured in Figure 4-3. We used the open-source code and pre-trained model provided by Siarohin et al., which is available on their Github repository[5]. The web-based system was implemented based on HTML and Javascript using the jsPsych Framework [1] with custom plugins for recording and uploading videos. Google Firebase was leveraged for hosting the website and storing the generated videos and recorded responses.

4.3.3 Procedure

The study was approved by the Committee on the Use of Humans as Experimental Subjects (COUHES), which serves as the institutional review board (IRB) for the Massachusetts Institute of Technology (MIT). Subjects were recruited via emails and public communication platforms on a voluntary basis. Each subject was required to sign a consent form, as well as submit a photograph of themselves and answer

[5] https://github.com/AliaksandrSiarohin/first-order-model

Figure 4-4: The study consisted of a pre-survey, a main study session comprising seven consecutive trials with three video conditions (None/Natural/Generated), and an exit survey. The first two trials featured the None condition to capture baseline measurements. The remaining trials comprised the Natural and Generated conditions, which were counterbalanced following a Latin square design.

initial questions in a pre-survey. Once their individualized materials were prepared, each subject received their custom link to the website to complete the study. At the beginning, participants were asked to imagine that they were entering a job interview, during which they would be asked a series of questions. For each question, they were told that they should try to answer the question as "clearly and confidently" as they could. They were also informed that their responses would be recorded.

The study followed a within-subject design. Within a single session, each subject was asked seven questions across seven distinct trials, which included one practice trial. For each trial, a subject was asked to (1) watch a video clip (20-25 seconds), (2) read and consider a question (15 seconds), (3) answer an interview question (1 minute), and (4) respond to survey questions (see 4-4). Participants watched one of three possible types of videos in each trial. They could either be given an abstract video with no speaker (**None**), a natural video of a speaker speaking confidently (**Natural**), or a synthesized video of him or herself speaking confidently (**Generated**). The *Natural* and *Generated* video clips each featured a single person answering an interview question in a professional manner (see Section 4.3.1). Videos were cropped into a square such that only speakers' heads and shoulders were visible (i.e. "talking-head videos").

Results from the first practice trial were discarded. In the following two **None** condition trials, an abstract video was shown to subjects. These were used to establish baselines for their speaking habits, that are not influenced by the proposed

interventions. Trials for the **Natural** and **Generated** conditions were then counter-balanced using a Latin-Square design to minimize learning effects. Stroop color tests [95] between trials were also used to help mitigate any influence from the previous trial. Since the confidence of an answer also depends on the questions, we opted to include two different categories of questions that were also counter-balanced: *factual* and *open-ended*. This was done to include a broad range of questions in a well-defined way. Examples of these were "What are your responsibilities in your current/most recent job" and "What is more important for success - talent or hard work?"

4.4 Analysis & Results

This is the first study we are aware of that investigates the perception of "*self-deepfakes*." We investigate the impact of our proposed approach based on participants' self-assessments reported through Likert-scales and open free-form comments.

4.4.1 Demographics

The study was completed by 28 participants (13 female). Participants' ages ranged from 20 to 60 years of age. They held many different occupations, including student, researcher, administrative assistant, registered nurse, lawyer, manager, product designer, and software developer.

4.4.2 Attitudes & Impressions

Oral Communication

Prior to experiencing the intervention, participants rated their level of confidence in impromptu speaking, their overall oral communication abilities, and how much they believed seeing confident speaking videos of themselves would boost their self-confidence in their ability to speak. The results (see Figure 4-5) suggest that the male

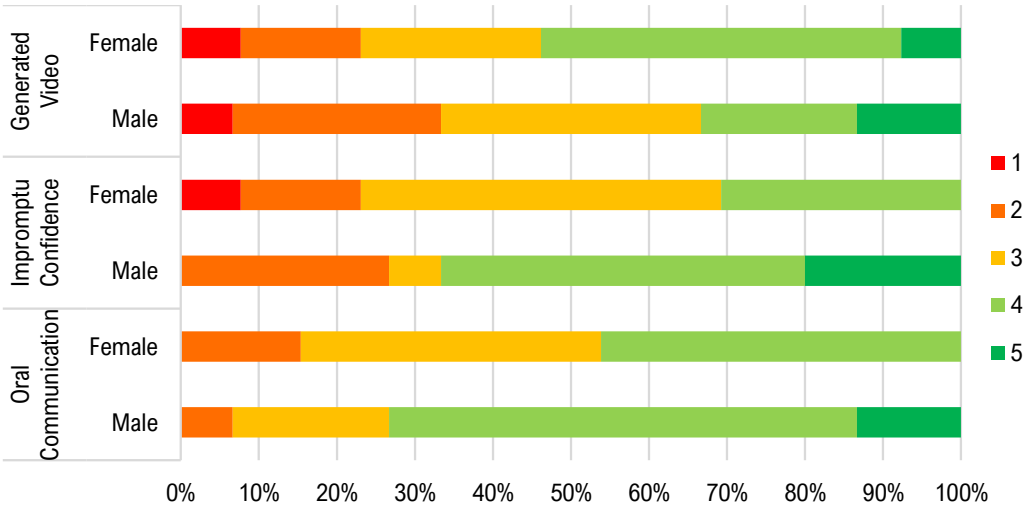


Figure 4-5: Participants were asked to rate themselves on a 5-point Likert scale regarding their oral communication skills (1 = Very Weak, 5 = Very Strong) and their confidence in impromptu speaking situations (1 = Very Unconfident, 5 = Very Confident). They were also asked to indicate their level of belief that seeing videos of themselves speaking well would boost their self-confidence in their ability to speak (1 = Not At All, 5 = Definitely).

subjects rated their oral communications skills higher than females and feel more confident in impromptu speaking situations. The females expressed greater optimism that synthesized videos of themselves could be helpful to boost their confidence.

Natural Speaker Videos

Participants' impressions of the Natural videos of example speakers varied. Categorizing their written feedback, 13 felt positive about them, 9 felt neutral, and 6 expressed that they did not like watching these videos. Those who appreciated them felt they were good role models for body-language (e.g. eye-contact, expressions, mannerisms), phrasing (e.g. confident vocabulary, argument structure), and vocal delivery (e.g. speech rate, clarity, enthusiasm, emphasis, smooth cadences). Conversely, some participants disliked them for their generic speech content, use of buzzwords, and attitude. Since the same videos were used as the driving videos for the generated videos, these points also applied to the Generated condition.

Generated Videos

Participants rated the video quality, as well as how much the person looked and sounded like them in the Generated videos. The numbers indicate a high degree of visual self-identification (median = 4) and high video quality (median = 4). As expected, "Sounds like me" ratings were much lower (median = 2) since audio was not synthesized and instead was taken directly from the professional speaker videos. Interestingly, visual resemblance was rated higher by the male subjects (median = 5) than the female subjects (median = 4), and the females rated auditory resemblance higher for Generated Videos (median = 4) than Natural videos (median = 3).

The experience of watching AI-generated videos of oneself elicited polarized responses. 14 participants felt positive about them, 10 felt negative, and 4 were neutral or unsure. Those who felt negatively mentioned that their general dislike of watching or listening to themselves in recordings contributed to their dislike of the experience. Some found it strange or discomforting, citing uncanny effects [71] due to glitches (e.g. choppiness, distorted face shapes) or general shock that their face could move with someone else's phrases, manner of voice, and behaviours. Of those who felt neutral towards the experience, two explained that after the initial shock, they found it interesting and believed that they could get used to it over time. Those with a positive outlook had varying comments. Some said it helped to have a "good example to follow" and "see how a good answer could look like." Another said they were satisfying, mentioning "...I can 'relate' because that was 'me."' Two mentioned that it encouraged being more animated in delivery and "less shy about [it]." Two said it instilled more confidence. Another two participants commented that the videos revealed possibilities, with one saying it was indirectly helpful "to acknowledge that there is a possibility for me to speak like that." Others found them to be reassuring, humorous and even relaxing.

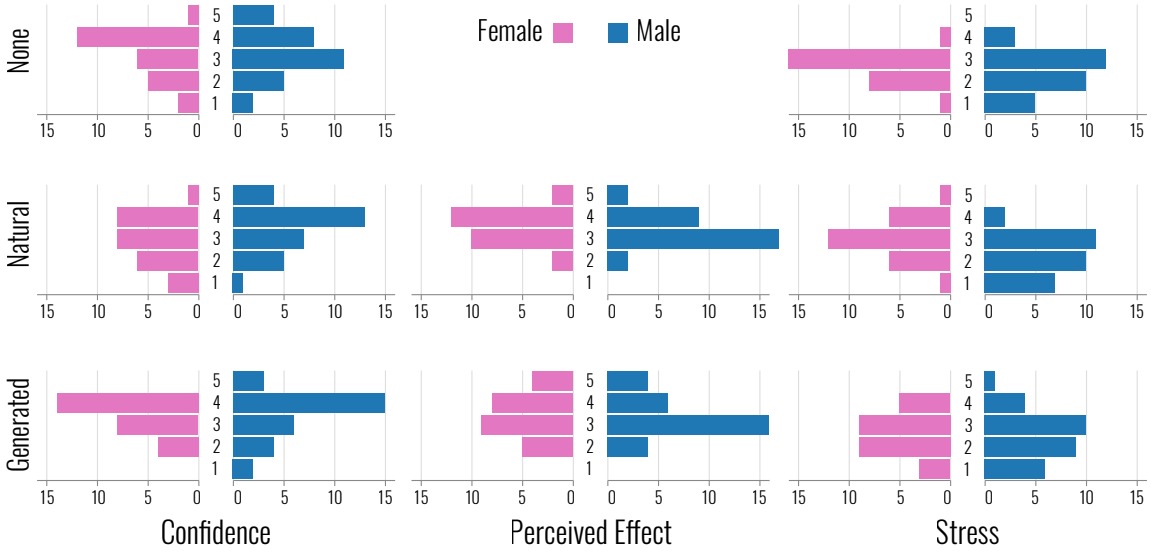


Figure 4-6: Summary of self-reported confidence, impact and stress in each condition. The bars show the response count per Likert scale. Depending on the question, 1 = Very low/negative, 5 = Very high/positive.

4.4.3 Subjective Effects

Participants were asked to self-report how they felt after answering each question. Their responses are shown in Figure 4-6. The impact the different conditions had on their feelings of confidence, performance, and stress was explored.

Perceived Feeling of Confidence

Participants were asked to rate how confidently they spoke when answering the interview question (1 = Very Unconfidently, 5 = Very Confidently). We observed unique distributions between male and female participants for self-reported confidence, as shown in Figure 4-7. Observing these distributions, it appears as though for the female participants, watching *Generated* videos prior to answering interview questions resulted in a boost of speaking confidence (i.e. after watching the synthetic video of themselves speaking confidently, they subsequently reported feeling they spoke more confidently when answering the interview question). In contrast, watching *Natural* videos appeared to have a negative effect on the females. For the male participants,

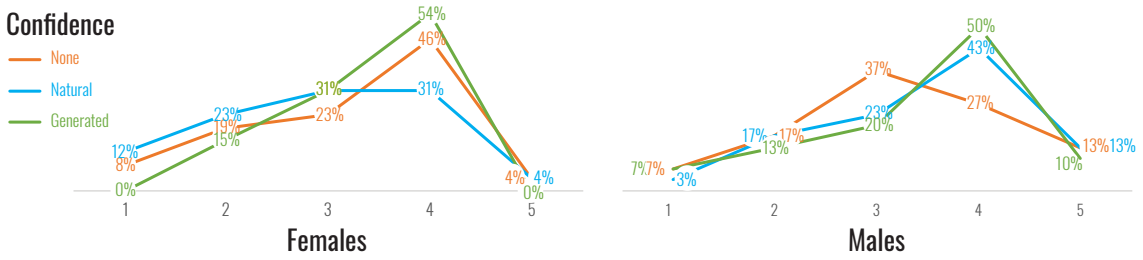


Figure 4-7: The distribution of ratings for confidence felt when answering questions in the different conditions, displayed by gender. Confidence was rated on a 5-point Likert scale per trial (1= Very Unconfidently, 5 = Very Confidently).

it seems as though males experienced a boost in confidence after watching either a *Natural* or *Generated* video. The median ratings for confidence reflect this as well (Males: None = 3, Natural = 4, Generated = 4, Females: None = 3.5, Natural = 3, Generated = 4). In open comments, one said watching generated videos of themselves instilled confidence in them, and two said they found it was reassuring.

Perceived Effect on Speaking Style

Participants were asked to rate whether they thought the video affected their speaking style positively or negatively (1 = Very Negatively, 5 = Very Positively). We observed a slight difference between the males and the females on this axis. Females considered the influence of Natural videos more positive (Median = 4) than Generated videos (Median = 3), while males were neutral towards both (Median = 3). Among those who felt the Natural videos had an effect, one person said it helped them recognize the faults in their own speech and triggered them to analyze how to improve. Two people felt intimidated and stated that it reminded them of their lack in their communication abilities. For Generated videos, some found them helpful as a positive role model and a means to visualize confident speaking. Interestingly one stated that it made them feel they were "speaking well," while another said it changed his belief that he could speak well: "I thought if a computer generated video of myself can speak confidently, then the real me can 100% speak confidently."

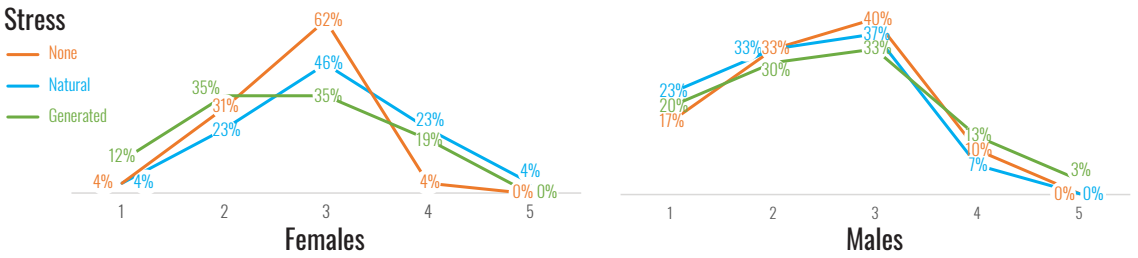


Figure 4-8: The distribution of ratings for stress felt when answering questions in the different conditions, displayed by gender. Stress was rated on a 5-point Likert scale per trial (1= Not At All Stressed, 5 = Extremely Stressed).

Perceived Level of Stress

Participants were asked to rate whether they felt stressed while answering the questions (1 = Not At All Stressed, 5 = Extremely Stressed). Answering interview questions is generally considered a stressful experience. We asked participants to rate how stressed they felt answering each question under each condition. We noticed that while the conditions did not appear to have much of an effect on the stress felt by the male participants, the distribution of responses of females shifted between conditions (see Figure 4-8). This suggests that watching any speaker video may be more stressful for females than none, but that Generated videos may cause less stress than Natural videos. We furthermore observed that male participants generally reported feeling less stressed than the females across each condition according to the median stress ratings (Males: None = 2.5, Natural = 2, Generated = 2.5, Females: None = 3, Natural = 3, Generated = 3).

4.5 Discussion & Limitations

The observations from the study reveal that the subjective impressions of both Natural speaker videos and Generated videos can vary greatly amongst participants. While Natural videos were overall perceived quite positively, the response to Gener-

ated videos was more polarized. Interestingly, our results also suggest that there may be nuances between genders. The female participants tended to appraise their oral communication abilities lower than males, and felt less confident in impromptu speaking situations. Although the females generally reported that watching others had a positive impact on their speaking style, they nevertheless tended to feel less confident in speaking after having seen these videos. The male participants rated themselves higher for impromptu speaking confidence and skill than females, and later reported feeling less stressed answering questions than the females. As the sample sizes were small and gender differences were not anticipated, it would be interesting to expand the sample population to more extensively investigate potential gender differences.

A general dislike of watching oneself was one reason driving negative opinions of the generated video experience. However, some participants also mentioned uncanny valley effects due to unusual visual artifacts and a mismatch of voices. Concerning the unusual visual artifacts, one person noted that their teeth were more prominent while another complained that the video was choppy. The first-order-model [88] has the advantage that it only requires one image of the target person, but it faces limitations (e.g. distortions to the face become apparent for extreme head angles) that can lead to some uncanny effects. As such, some impressions of generated videos gathered in this study are specific to the outcomes of this model, and future models with better quality output may elicit different reactions and impressions. Generated videos were limited to using the voice of the speaker in the driving video. Some participants commented on the uncanny effect that came from a mismatch in voices in their generated videos. However, especially amongst females, the ratings for auditory resemblance for generated videos were higher for natural voices despite the fact that the same voices were used. This suggests that greater visual likeness can for some people increase the perception that a voice is one's own. Since state-of-art open-source voice cloning techniques are still not capable of producing highly realistic results (i.e. with a natural prosody) without extensive input and training, we did not incorporate

this feature into the generated videos. In the future, incorporating realistic generated voices would improve impressions by reducing the uncanny-valley effect.

In this study, we analyzed the data using qualitative feedback from participants. As an extension to this work, it would be interesting to use human raters to analyze the responses of participants according to various speech factors such as conciseness and pacing, as per findings from the online survey (see Section 3). Restrictions from COUHES made it difficult to involve external experts in the coding and analysis of the data. Alternatively, it may be possible to use automated methods to quantitatively analyze the recordings. We explored the use of speech analysis software such as PRAAT [13] to analyze voice features in the recorded audio. However, careful considerations must be made into the scope of such analyses. Previous work studying the relation between acoustic features and confidence (refer to Section 2.3.1) was based on single statements. Responses collected in this study were much longer (1 minute) and comprised multiple sentences. We can consider that an appropriate, competent, and confident response may comprise a mixture of sentences stated with various levels of confidence. Furthermore, confidence comprises many factors beyond vocal delivery, such as the structure and conciseness of a response (refer to Chapter 3). Therefore, we render the results of the automated speech analysis as inconclusive and see it as an open area for research to determine how longer responses can be appropriately analyzed with respect to these features.

Speaking confidence is affected by many variables. As such, there are multiple parameters for the design of the study that can be tuned, such as the context, the content, and the speaking partner. We created a web-based experience in which people answered interview questions on camera instead of speaking with a live interviewer. This configuration provided a distinct experience, with many factors that may have both added to and reduced the stress of the experience compared to a standard interview situation. It would therefore be interesting to extend this work by repeating the procedure with a live interviewer that could give verbal and non-verbal feedback.

Nevertheless, we opted for this study design and setup as it was scalable and would offer a consistent experience between participants.

The comments and reactions of participants point to potential areas for further investigation. In this study, participants were tasked to speak immediately after they were presented with generated videos of themselves. However, some suggested that the ability to acclimatize or familiarize themselves with their generated representation would be beneficial. As such, it may be helpful to conduct a follow-up study in which people are given time to mentally connect with their virtual representations, in line with existing strategies from research in VR and avatars (see Section 2.2.1).

In hindsight, we suspect that the brevity and the stressful context of the study scenario (i.e. the fact that it was an interview situation and were informed that recordings of them would be reviewed by experts), may have triggered participants to resort to their usual habits rather than actively experiment with or embrace changes to their speaking style. In addition to this, some participants commented that they did not like the speaking style of the example speakers. Building on this, we believe our approach could be better leveraged in a private communication practice system, where people can control the generation of their own "*deepfakes*," use them to experiment with their delivery, and finally reflect on and consciously adjust their speaking styles in response to their observations. The advantages of this would be two-fold. First, giving people a role in the creation of these videos could help to reduce the shock, negativity, or resistance towards seeing them, since they could then choose speakers that they like as well as build familiarity with their generated representation. Second, it would give them a safe space to learn through play with their speech delivery, and even try on different personalities, without any negative consequences. As one participant stated: "Very intrigued and surprised to see myself in different manners or almost different personality... (I) feel somehow more competitive with myself..."

4.6 Summarizing Remarks

We performed a novel investigation into people's impressions of seeing personalized AI-generated videos of themselves in the context of communication. In our user-study with 28 participants, we compared the impressions and relative impact of watching different videos as a stimulus, prior to asking them to answer interview questions. Participants were presented with videos with no speaker, with another person as a confident speaker, and with an AI-generated version of themselves speaking confidently. Our study provided preliminary evidence that some subjects may find Natural and Generated videos to be helpful as role models. Furthermore, they enabled some subjects to reflect on how they could improve their own performance. Generated videos had a more polarizing effect on the participants than natural ones, but some of their comments suggested that increased familiarity may change their perspectives. Our results also suggest that genders may be impacted differently by this approach in the context of communication. Overall, we see that this experiment opens up a new avenue for exploration of self-help opportunities and of constructive applications for synthesized deepfake videos.

For future work, we would like to experiment with designing and implementing a system that gives participants the ability to experiment with the generation of their own "*deepfakes*." This would serve to give them greater time to familiarize themselves with their synthesized representations. It would also allow them to actively choose their own role-models upon which their synthesized representations would be based. This would likely strengthen the impact of the system by making the generated videos more appealing to observe and mimic. As a technical next step, we would like to use AI techniques to synthesize people's voices in addition to their videos. While only some participants expressed discomfort from using the voices of others, we suspect this could also help strengthen the impact of the overall experience.

Chapter 5

User Study 2: Exploring Real-Time Camera Filters for Creativity

Real-time virtual camera filters have become commonplace in the landscape of online communication tools. Largely designed for entertainment purposes, they enable people to customize and change their appearance in fun ways for photos and videos. In this chapter, we present a novel investigation into using real-time virtual camera filters for the constructive purpose of altering people's perception of themselves to improve their cognitive abilities. We investigate this concept in the context of creativity and conduct a user study where participants are asked to apply different camera filters in a video call before engaging in creative tasks. We explain the details and findings of this study and summarize the potential real-time camera filters can play in assisting with cognitive augmentation and self-improvement.

5.1 Motivation

Beyond connecting individuals online, many video conferencing platforms offer the unique ability to digitally manipulate visuals during conversations. For instance, the

look of one's environment can be changed with the use of virtual backgrounds. Additionally, a growing number of platforms provide real-time camera filters (e.g. Zoom [112] and Snap Camera [90]), which present an accessible avenue for manipulating one's appearance. This is particularly interesting as they not only enable us to alter how others perceive us, but also allow us to change how we perceive ourselves.

Studies have shown that virtually embodying someone who we perceive as having greater cognitive abilities than our own may have a positive impact on our own capabilities. Men who had low self-esteem and who embodied Einstein demonstrated an improvement in their logical thinking versus those who embodied a standard male avatar [8], and engineering students who embodied inventor avatars performed better in group brainstorming activities than when no avatar or non-inventor avatars were used [40]. While virtual reality (VR) technology has enabled these fascinating outcomes, the technology is still not widely accessible and used. As such, we are motivated to explore the potential for increasingly pervasive real-time camera filters to achieve a similar outcome. By applying such filters and showing them in a self-view, we explore the possibility to alter people's self-perceptions in order to improve their cognition. While camera filter effects on a computer are limited to the size of a display in contrast to VR head-mounted displays (HMDs), filters are a highly accessible and scalable technology. Furthermore, people can engage in many activities including brainstorming without needing to divert their gaze from the screen, making it a viable technological approach for manipulating self-image and cognition. However, it is still unknown to what degree real-time camera filters can support a sense of embodiment, particularly when used on their own. Therefore, we conduct a study to examine the potential impact real-time camera filters can have on self-perception, and extend our exploration to investigate its influence on cognition in the realm of creative thinking.

5.2 Concept

In this chapter, we present the concept of constructively leveraging real-time camera filter technology as a means to alter people's digital self-image during video calls and consequently positively influence their cognitive capabilities. Their use could be integrated into the video-conferencing experience by simply applying them during the course of a meeting. Prior work in VR suggests that embodying a different person with superior cognitive capabilities such as an inventor [40] or Einstein [8] can boost one's cognitive capabilities (refer to Section 2.2.1). We see the potential for real-time filters to serve as a highly accessible means to similarly alter how people see themselves to achieve a similar effect.

We contribute a preliminary investigation into the potential impact and use of real-time filters in the context of creativity. Self-view is a commonly used feature in many online communication tools (refer to results of the survey on online communication in Chapter 3) that gives people the opportunity to see themselves in an interaction with others. Motivated by the prospect of using this as an input channel, we conducted a remote user study with 21 participants. As pictured in Figure 5-1 we compared participants' experience of a *no-filter* condition (A) to real-time camera

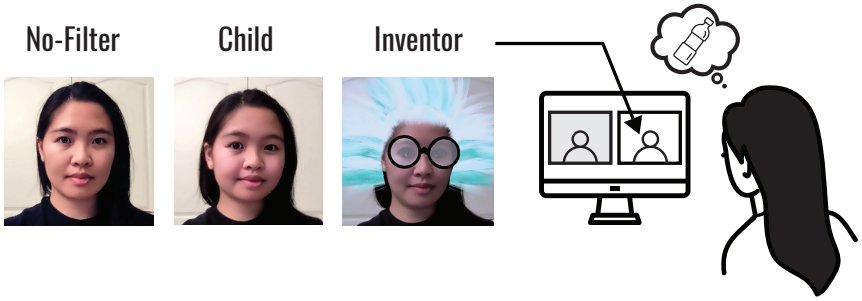


Figure 5-1: In a user study, we explore the impact of different real-time virtual camera filter conditions on people's moods, feelings of embodiment, and performance in creative tasks. Participants were asked to apply camera conditions to be visible in their self-view during a 1-to-1 video call. The filter conditions were a no-filter condition, a child filter condition, and an inventor filter condition.

filter conditions, *inventor (I)* and *child (C)*. The inventor and child filters were selected since inventors and children are generally believed to be highly creative thinkers [40, 21]. Subjects' perceptions of embodiment and mood were captured to investigate what underlying psychological processes and mechanisms may underpin filter effects. Divergent thinking tasks were conducted to investigate the potential impacts filters may have on creative cognitive ability. We summarize our findings and discuss ideas for future research in this direction.

5.3 User Study

We conducted an online experiment over Zoom to explore the effects of real-time filters on creativity. Cognitive and affective factors, such as mood and embodiment, were examined to understand potential underlying mechanisms and processes that can underpin effects. While we considered investigating filters in group settings, as they may yield stronger effects, we decided to first study the applicability of filters on oneself to understand their effects on self-image manipulation. As we focused exclusively on self-image manipulation, our study also excluded the use of virtual backgrounds, which could create the environmental illusion of being in a classroom or science lab may extend the embodiment illusion and have additional cognitive effects. It would therefore also be an interesting dimension to explore in future work.

5.3.1 Conditions

Each participant experienced three filter conditions: **a no-filter, adult condition (A), an inventor[1] filter condition (I), and a child[2] filter condition (C)**. The

[1]Mad Scientist by Charles Hamblen:
`https://www.snapchat.com/unlock/?type=SNAPCODE&uuid=f341b6ab08254b8bb0e46ffc02409280`

[2]Baby by Snap Inc.:
`https://www.snapchat.com/unlock/?type=SNAPCODE&uuid=69a3ae3fe3bb4007ba514afda7d3a97d`

no-filter condition presented a standard unmodified camera stream of the person, which served as a baseline condition. The inventor filter changed people's appearances by enlarging their eyes, adding spiky white hair with highlights, and adding goggles over the face. The child filter changed the proportions of people's faces to be smaller with larger cheeks and teeth, as well as smoothened skin and removed adult facial hair (i.e. beards/mustaches). Comparing the inventor and child filters, the inventor filter appeared more like a mask by superimposing features, whereas the child filter did more to morph facial features. An example of these conditions are pictured in Figure 5-1. The filters were selected since they do not completely conceal a person's facial features. Rather, they manipulate facial features in a way that the person is still recognizable. The choice of filters corresponded with the theme of creativity, wherein inventors and children are generally understood to be creative thinkers [40, 21].

5.3.2 Experimental Set-Up & Apparatus

The study was conducted remotely. Consent forms and pre-surveys were distributed via Docusign and Google Forms. For the main study, each subject was required to have Snap Camera [90] and Zoom [112] installed. Participants connected with the study coordinator via a password-protected Zoom call, and selected Snap Camera as their camera input. On the call, subjects were asked to run both applications simultaneously, keep their video cameras on throughout the study, and when instructed, enable or disable the specified camera filters. During the study, filters were only applied by the participants and not by the study coordinator. Under Zoom's video settings, they were asked to have "Mirror my video" and "Adjust for low light (Auto)" checked. They also enabled "Gallery View", which made their self-view the same size as the study coordinator's video stream on their Zoom interface. If necessary, participants were asked to adapt their physical set-ups (e.g. adjust lights for flat lighting across the face, adjust webcam to center the face in the camera stream).

Each session was recorded. Links to the surveys, created using Qualtrics [85], and access URLs for the respective camera filters were provided step-by-step via the chat functionality of Zoom. Subjects were instructed to perform all study activities on a single monitor, and have only a single window open at any one time filling most of their screen (i.e. either the Zoom window or web browser). The study coordinator used a plain white virtual background to maintain consistency between participants. Participants on the other hand were instructed to disable their virtual backgrounds to ensure (1) that their camera filters worked smoothly and (2) that their onscreen backgrounds matched their true physical surroundings. The study coordinator's video camera remained on at all times (except during the Survey phases), and the coordinator was available at any moment to answer any questions. Participants were generally not given special instructions on how or where to look during the call in order to mirror normal video-conferencing conditions (except for a portion of the orientation phase and a 20-second period prior to the survey phase, where they were instructed to look at their self-view).

5.3.3 Standardized Tests & Response Variables

We issued surveys based on standardized tests for mood, embodiment, and creativity, which are explained briefly below.

Mini-IPIP (Personality): The International Personality Item Pool (Mini-IPIP) is a 20 question personality scale. It is a short version of the original IPIP that produces the Big Five personality traits (i.e. extroversion, agreeableness, conscientiousness, emotional stability, and intellect/imagination).

IAT (Mental Associations): The Implicit Association Test [39] is a tool commonly used in psychological research to measure mental associations between two target concepts (e.g. flower, insect) with an attribute (e.g. pleasant-unpleasant). It

tasks people with categorizing stimuli according to these targets and attributes and records reaction times. Resulting dScores are based on average time differentials between the two pre-defined associations. In this study, we focused on the targets *self* and *other* against the attribute of *youth* (i.e. *child-adult*). Using information from the pre-survey (i.e. full name, hometown, home country, native language, occupation) and black-and-white portrait photos of adults and children, we generated personalized self-other child-adult IATs using iatgen [19], which we issued via Qualtrics [85]. In interpreting the results in our study, positive dScores indicate stronger associations for Me-Child relative to Me-Adult, negative scores suggest the opposite, and a zero dScore indicates no bias [19]. While it is a popular tool, we note that the reliability and validity of these tests have been questioned[1].

I-PANAS-SF (Mood): The International Positive and Negative Affect Schedule short form (I-PANAS-SF) [56] is a survey in which respondents rate the intensity of emotions experienced along a 5-point Likert scale (1 = Very slightly or not at all, 2 = A little, 3 = Moderately, 4 = Quite a bit, 5 = Extremely). It spans 10 different emotions, classified as positive or negative emotions. Positive emotions include: *alert, inspired, determined, attentive, and active.* Negative emotions include: *upset, hostile, ashamed, nervous, and afraid.* From these ratings, a net *positive* and *negative* affect score can be calculated (min = 5, max = 25) by summing the corresponding emotions.

Standardized Embodiment Questionnaire: This questionnaire was devised by Gonzalez-Franco and Peck [36] to capture the extent to which the embodiment illusion is experienced amongst participants. The embodiment illusion refers to the feeling that one's body has been substituted by the avatar body (refer to the related work Section 2.2.1). In our study, participants answered 17 7-point Likert-scale questions based on [36]. Responses ranged from *strongly disagree* (-3) to *strongly agree* (+3).

[1] `http://www.hcdi.net/reliability-and-validity-of-implicit-association-test/`

We explored a subset of the original variables since *response to external stimuli* did not apply. From [36], we asked questions on *body ownership* (1-5), *agency* (6-9), *tactile sensations* (10-11), *location* (14-16) and *appearance* (17-20). Some of the wordings for questions were slightly adapted to match the context of this study. Higher scores for each axis indicate greater degrees of embodiment. *Body ownership* is how much a person feels they own the avatar body and *agency* is the feeling of control over it. The other factors either enhance or detract from the embodiment illusion [36].

AUT - Alternate Uses Task (Creativity): The Alternate Uses Task (AUT) [42] is a test that involves having the test taker brainstorm as many unique and unusual uses for a given object as one can within a given time limit. It is used to measure creative divergent thinking capabilities (refer to Section 2.4.1). In our study, participants repeated this task based on the following prompts: *pizza box, plastic water bottle, broom, paper clip, plastic fork, and rubber band.* For each object, they were instructed to say their ideas aloud. Two objects were given per condition/trial. The study coordinator would nod in response to each idea as an indication to the participant to continue and move on to the next idea.

VGT - Verb Generation Task (Creativity): The Verb Generation Task (VGT) [83] is an approach for measuring a person's creative divergent thinking capabilities (refer to Section 2.4.1). In this test, participants provide a single verb in response to a given noun (e.g. for the word 'scissors' one might say 'cut' or 'throw'). The more distantly related the words are, as determined by Latent Semantic Analysis (LSA), the more creative the response. In our study, participants completed the task verbally and were prompted to be as creative as they can for each answer. Three unique lists of 16 nouns were asked per condition/trial. The word lists were derived from the original list of nouns published in [83], such that each list contained words with equal frequency values. The order of the words was randomized per participant.

5.3.4 Procedure

The study was approved by the Committee on the Use of Humans as Experimental Subjects (COUHES), which serves as the institutional review board (IRB) for the Massachusetts Institute of Technology (MIT). People were recruited via emails and public communication platforms, and each subject was paid with a $25 Amazon gift card as a thank-you for their participation. Each participant was required to sign a consent form, as well as answer a pre-survey comprising basic demographic questions and personality questions from the Mini-IPIP. The study took approximately 75 min. for completion per participant.

The study followed a within-subject repeated measures design. Participants first completed a self-other child-adult IAT to capture a baseline measurement. They then experienced three counterbalanced conditions sequentially within one session. The overall study protocol is pictured in Figure 5-2. Each condition consisted of a *setup/orientation*, *creativity assessment*, and *survey* phase. To begin, subjects were asked to apply one of the three filter options: *No-Filter* (A), *Inventor* (I), or *Child* (C). To orient subjects with the respective filters, the study-coordinator both announced and performed a predefined set of physical actions for the subject to copy (e.g. "shrug your shoulders", "touch the top of your head", "raise your eyebrows up and down"). This process was motivated by the *Enfacement Illusion* [103] that shows that the perception of one's own self-image changes with tactile stimulation, and was also motivated by VR studies that use visual-motor synchrony to enhance feelings of body-ownership [36]. An additional 20 seconds were given for each subject

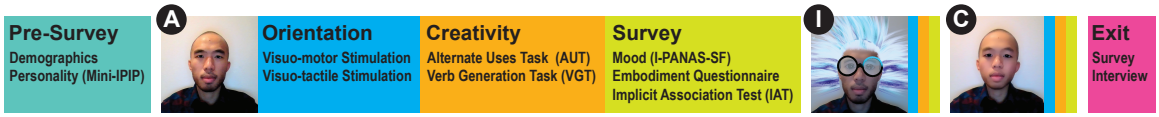


Figure 5-2: Each participant experienced all three filter conditions in three separate trials. From left to right: no-filter/adult (condition A), the inventor filter (condition I), and the child filter (condition C). Conditions were counter-balanced between participants.

to continue to familiarize themselves with their camera-stream. During the *creativity assessment* phase, subjects completed a verbal Alternate Uses Task (AUT) [42] for two distinct objects (e.g. broom, water bottle). For each item, they had two minutes to think of as many unique and unusual uses for this object as they could. This was followed by a Verb Generation Task (VGT) [83], where they were cued to say a verb aloud that could be associated with a provided noun. Objects for the AUT and noun sets for the VGT were altered between conditions. A 20-second period of time was then given to participants to look at their camera image, to provide them with an opportunity to quietly reexamine their appearance before proceeding to the survey. The trial concluded with the survey that comprised questions about their brainstorming experience, mood (I-PANAS-SF) [56], embodiment (following the standardized questions proposed in [36], and another IAT [39, 19]. Subjects were invited to take a short break before proceeding to the next trial. At the conclusion of all three trials, subjects completed an exit-survey and were asked to openly provide comments about their overall experience and impressions of the activities and the different filters.

5.4 Analysis & Results

We summarize our findings from the study regarding how real-time virtual camera filters affect mood, embodiment, and creativity.

5.4.1 Demographics

The study was completed by 21 participants (14 female) aged 18 to 31. They comprised 14 undergrad and graduate students, 3 engineers, an entrepreneur, a project manager, and a program director. One was unemployed. 18 were native English speakers. They were asked to rate their familiarity with using virtual camera filters

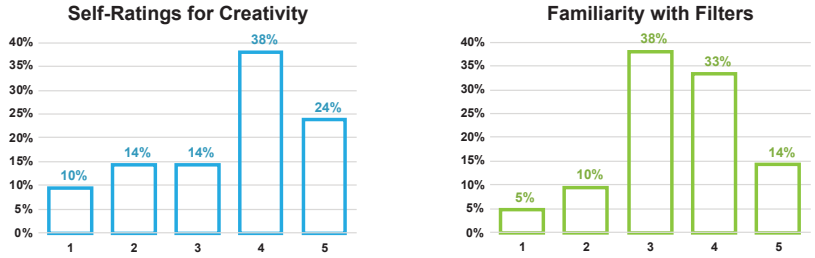


Figure 5-3: Participants generally appraised themselves as being creative, and most were at least somewhat familiar with camera filters.

on a 5-point Likert scale (1 = Not at all familiar, 5 = Very familiar). Most participants reported believing they were creative in their pre-survey (" I am creative.", 1 = Very Inaccurate, 5 = Very Accurate) with 62% responding with a rating of 4 or 5 (see Figure 5-3, left). Participants were mostly already familiar with the use of real-time camera filters, with 85% of the participants reporting being *somewhat familiar*(3) to *very familiar*(5) with them (see to Figure 5-3, right).

5.4.2 Personality

Participants were assessed for their personality traits using the Mini-IPIP test that was integrated into the pre-survey. We observed that our participants are generally quite Agreeable. One notable trait of our subject population is that there is a particularly large score span for Conscientiousness and Intellect-Imagination as can be seen in Figure 5-4.

5.4.3 Filter Impressions & Beliefs

Participants were also asked to report their impressions of the filter conditions (see Figure 5-5). They generally reported liking the effects of the inventor filter more than the child filter. People were also asked how old they felt they looked. The child filter successfully made people feel younger, however, age perception varied much more

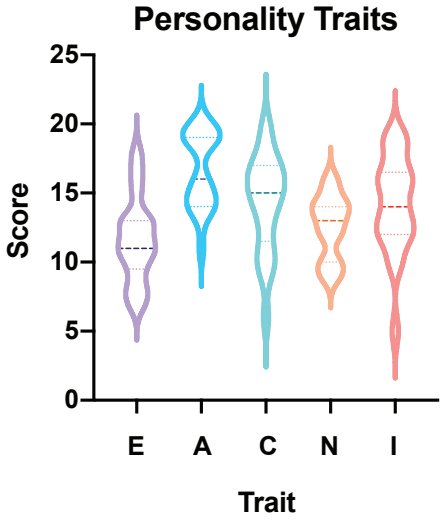


Figure 5-4: Personality trait scores of the participants, including Extroversion (E), Agreeableness (A), Conscientiousness (C), Neuroticism (N) and Intellect-Imagination (I).

greatly with the inventor filter applied. The mean perceived ages were 24.8, 35.0, and 8.0 for the Adult, Inventor, and Child conditions respectively, while on average participants were 24.3 years of age.

Participants were asked to indicate on a 7-point Likert scale "If a creativity test were to show age differences, who do you think would perform better?" (1 = Children Perform Better, 7 = Adults Perform Better). The median response for all participants was 2. Participants were also asked to rate on a 7-point Likert scale to what extent they think that a stereotype exists in the U.S. that children are more creative compared to adults (1 = No Stereotype, 7 = Very Strong Stereotype). The median response for all participants was 6. As such, it could be said that our participants generally believed that if age is a factor for creativity, children would outperform adults. They also believe there is a strong stereotype that children are more creative than adults.

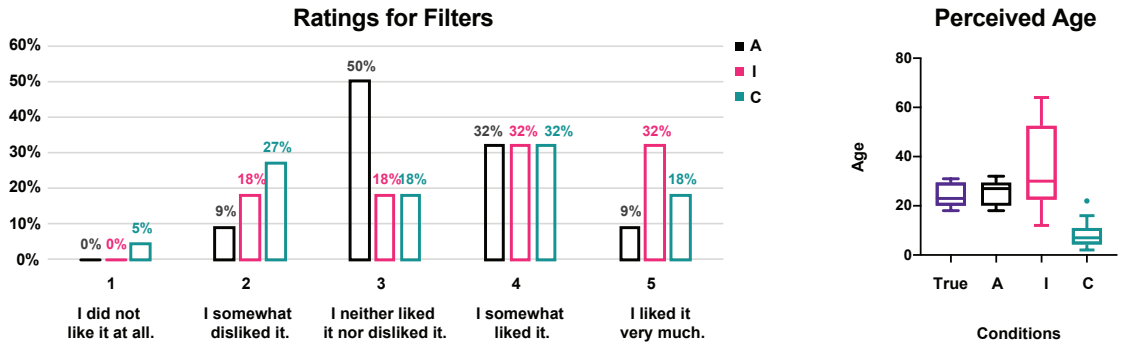


Figure 5-5: Participant impressions of the different filter conditions. Left: participants rated on a 5-point Likert Scale how much they liked the filter. Generally, the inventor filter was favoured over the child filter. Right: the child filter successfully made people look younger to themselves, while the inventor filter had a more ambiguous effect on people's age perception.

5.4.4 Embodiment

The total embodiment score, computed using arithmetic addition and weights proposed in [36], was on average highest for condition A, followed by C and I (see Figure 5-6). A Friedman test was carried out to compare the total embodiment scores for the 3 conditions ($\chi 2(2) = 16.65$, $p < 0.001$). Wilcoxon matched-pairs signed-rank tests yielded significant differences in the medians across groups I-A ($Z=-3.74$, $p<0.001$) and C-A ($Z=-3.52$, $p<0.001$) on this axis. We note that these results must be interpreted cautiously due to the small sample size of participants. These scores suggest that the participants felt a higher total sense of embodiment in the adult condition, which corresponds with no filter. This is not surprising, since no particular distortions are made in this condition that can contradict people's understandings of their real physical bodies. A Wilcoxon matched-pairs signed-rank test did not detect a significant difference between the medians for I and C regarding total embodiment, although C is observed to score slightly higher than I ($Z = 1.11$, $p = 0.25$). This may be an indication that more realistic filter effects may be able to draw higher feelings of embodiment.

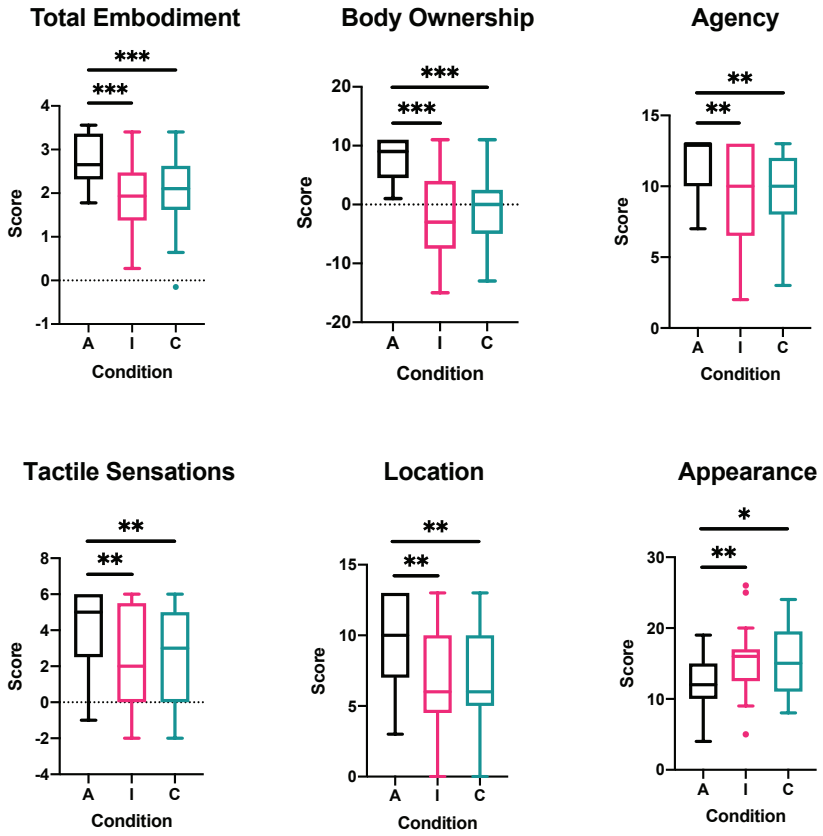


Figure 5-6: No-filter condition A significantly differed from the filter conditions I and C in total embodiment and the related factors from [36]. I and C were not found to be significantly different from each other on these axes. Note: $p<0.001$ = ***, $p<0.01$ = **, $p<0.05$ =*. The box and whisker plots were produced with the Tukey method, where points greater than the 75th percentile plus 1.5 times IQR or lower than the 25th percentile minus 1.5 times the IQR are plotted as individual points.

With regards to the individual embodiment factors, the scores for the *no-filter* condition A are observed to be on average higher than those for *filter* conditions I and C. A Friedman test detected a significant effect of filter condition on *body ownership* scores ($\chi2(2) = 28.92$, $p < 0.001$), and Wilcoxon matched-pairs signed-rank tests yielded significant differences in the medians across groups I-A (Z=-3.30, p<0.001) and C-A (Z=-3.30,in p<0.001). Similarly, Friedman tests were conducted for *agency* ($\chi2(2) = 11.11$, p<0.01), *tactile sensations* ($\chi2(2) = 12.25$, $p < 0.01$), *location* ($\chi2(2) = 10.48$, $p < 0.01$) and *appearance* ($\chi2(2) = 8.43$, $p < 0.05$), with each test detecting that there was a significant effect of filter condition on the respective scores. Wilcoxon matched-pairs signed-rank tests for these factors yielded significant differences in the medians across groups I-A: *agency* (Z=-1.98, p<0.01), *tactile sensations* (Z=-1.56, p<0.01), *location* (Z=-2.21, p<0.01), and *appearance* (Z=2.47, p<0.01). Wilcoxon matched-pairs signed-rank tests also yielded significant differences in the medians across groups C-A: *agency* (Z=-1.77, p<0.01), *tactile sensations* (Z=-1.74, p<0.01), *location* (Z=-2.61, p<0.01), and *appearance* (Z=1.89, p<0.05). These results are pictured in Figure 5-6. We note that appearance scores for condition A are lower. This is anticipated since the questions were attuned to a situation where the avatar can be logically perceived as a separate entity (e.g. Q17. "It felt as if my real body was turning into an avatar body" would yield a lower score for an accurate camera image over a less realistic filter condition). Overall, these study observations suggest that real-time camera filters may not be able to induce as high levels of embodiment that would be perceived from one's unmodified video image (however this would be difficult for any such filter to attain). We can also note that feelings of embodiment may differ the most between filters with respect to the variable of *body ownership*.

5.4.5 Mood

Prior research indicates that positive emotions are linked to higher performance in creativity [29]. To understand the potential emotional impact of the different filter conditions, study participants were asked to complete the International Positive and Negative Affect Schedule Short Form (I-PANAS-SF) [56], described in Section 5.3.3. The conditions did not differ greatly for the net positive affect score (means: A=17, I=17.5, C=15.5) and negative affect score (means: A=7.4, I=7.3, C=8.3). A subset of the emotion ratings is shown in Figure 5-7. C made participants feel more *Ashamed*, with only 52% reporting 1 (Very slightly or not at all) vs. 76% for A and 67% for I. They also felt less *Alert* (rating $>$ 4: A=52%, I=62%, C=43%), *Determined* (rating $>$ 4: A=62%, I=67%, C=48%), *Active* (rating $>$ 4: A=62%, I=71%, C=52%), and *Attentive* (rating$>$4: A=67%, I=62%, C=57%). In the child condition, participants felt less inspired as 29% reported 1 in contrast to only 10% for A and I. Generally, similar emotional tendencies exist between A and I, where they tend to score higher on the positive emotions but lower for the negative emotion of shame.

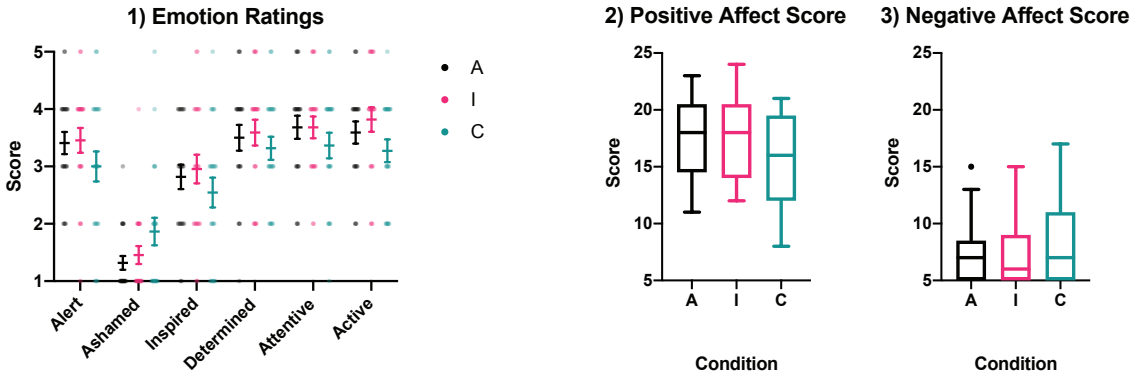


Figure 5-7: Results from the I-PANAS-SF mood assessment. The plot includes the mean and SEM bars. These were included even though the plot contains Likert data, to showcase the subtle differences of the child condition on many metrics. 1) Subset of emotion ratings by the study participants, 2,3) the net positive and negative affect scores per condition. The box and whisker plots are produced using the Tukey method. In general, it appears as if condition A and I trigger similar emotions, whereas C triggers different tendencies.

C surprisingly had a negative effect on emotions for participants. Overall, these responses indicate that real-time camera filters may trigger different emotions in some, but more research is needed to understand the underlying reasons.

5.4.6 Creativity

Subjects completed two activities pertaining to creativity: the Alternate Uses Task (AUT) and the Verb-Generation Task (VGT). Prabhakaran et al. [83] proposed that semantic distances between pairs of text positively correlate with creativity. AUT results were analyzed using SemDis [9], a novel web-based tool that computes semantic distances between brainstorming prompts and corresponding ideas to assess creativity. Higher SemDis scores (i.e. 'SemDis_MEAN' metric) between two texts indicate they are more distantly associated concepts and thus indicate more creative solutions. SemDis cleans the data by removing special characters and numbers as well as filler and stop words. A distinct score is calculated per prompt and response. As shown in Figure 5-8 (left) a score per trial was calculated by taking an average of the

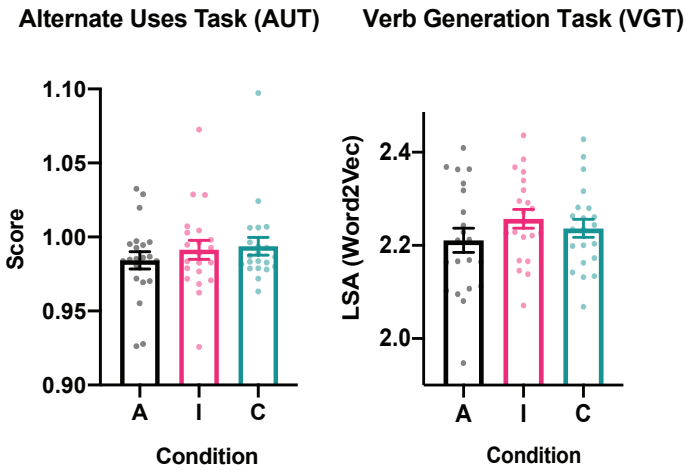


Figure 5-8: Participants performance on the AUT and VGT. Semantic distance scores for both tasks were slightly higher for the filter conditions I and C. The plots show the mean with Standard Error of the Mean (SEM) bars.

scores per item, and then the average of these. While Fluency (i.e. number of ideas generated) is typically of interest for the AUT, there were no considerable differences between conditions for this since the activity was tightly time-capped. For the VGT, semantic distances between each noun-verb pair were calculated using latent semantic analysis (LSA) based on the gensim Word2Vec model[2] (16 words per trial, 48 pairs in total). We averaged the 16 scores from each trial to create a single score representing a participant's performance in the corresponding filter condition. A plot of these scores is shown in Figure 5-8 (right). As the order of conditions was counter-balanced, our preliminary results from these tests suggest that there may be a slight tendency for higher creativity in the I and C conditions which is very encouraging to see; however, more participants would be needed for more conclusive results.

5.4.7 Implicit Associations

We conducted the IAT to investigate whether the child filter affects people's self-association with being child-like. Unexpectedly, averaged over all samples, (see Figure 5-9, 1), condition C showed a negative score (higher self-association with adult) while A and I showed positive dScores. Investigating this, we noticed that this tendency of having higher values for A than for C was not consistent across participants. In fact, it was only the case for 10 subjects (*Insusceptible*), while the trend was reversed (higher association with child in the child condition, as expected) for the other half (*Susceptible*, 11). Figure 5-9 (2,3) shows these opposing trends for A and C between these groups. The baseline dScores (*Pre*) for the Susceptible group shows an initial stronger self-association with adult (expected for our adult subjects), whereas is almost neutral for the Insusceptible group. Conscientiousness levels differed between the groups (Median Suscept. = 12, Insuscept. = 16) and *Intellect-Imagination* (Median Suscept. = 12, Insuscept. = 16), as shown in Figure 5-9 (4,5). Mann-Whitney

[2] `https://radimrehurek.com/gensim/models/word2vec.html`

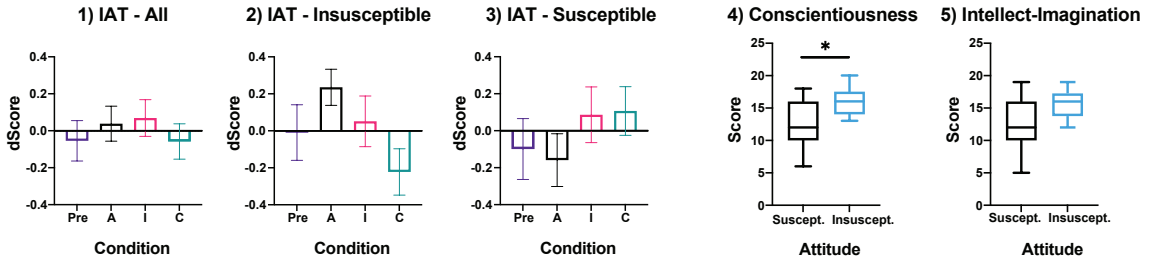


Figure 5-9: 1) Bar chart representing the mean and SE of IAT Scores by condition for all participants. Pre represents a baseline that was created before any condition was started. 2,3) IAT Scores by categorization of participants into two groups that are either "susceptible" or "insusceptible" to the filters. 4,5) These groups show a considerable correlation with Conscientiousness and Intellect-Imagination personality scores [31].

tests found (U = 27, p<0.05, two-tailed) for *Conscientiousness*, and (U = 30, p = .078, two-tailed) for *Intellect-Imagination*. Larger sample sizes would be needed to determine a significant difference. Generally, these observations suggest that there may be different groups of people with different reactions to such visual manipulations.

We were also interested in the potential relationship between our susceptibility (S = Susceptible, I = Insusceptible) categorization and the creativity scores from the AUT and VGT. Splitting the AUT and VGT scores based on this categorization, we observed slight but notable differences between these two groups (see Figure 5-10). A surprising result was that the two groups had different tendencies in the AUT and VGT. Wherein the Insusceptible group on average scores higher than the Susceptible group on the AUT, it was exactly the opposite for the VGT. As both tests are designed to measure creativity this is an unexpected result that warrants further investigation.

5.5 Discussion & Limitations

Participants were asked to share their comments on the experience at the conclusion of the study. Some participants reported that they felt the filters impacted their levels of inhibition. For some, they felt that the filters acted as "masks" and reduced their

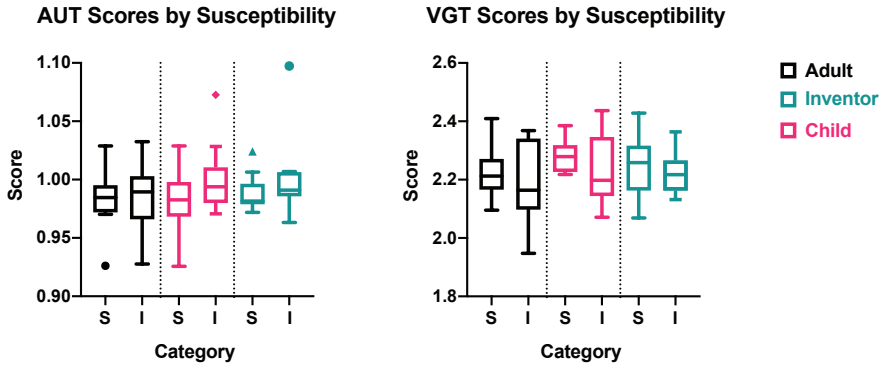


Figure 5-10: Creativity scores from the AUT and VGT according to the categorization of susceptible (S) and insusceptible (I). The box and whisker diagrams are plotted with the Tukey method.

feelings of self-consciousness. Consequently, they felt more liberated from judgement and found it easier to produce more creative ideas. However, others felt more self-conscious, particularly with the child filter, with a portion believing that it detracted from their creative output. In future work, it would be interesting to determine whether people compare themselves with their conversation partners. In the study, we refrained from having the study-coordinator apply the same filter since we wanted to exclusively test the impact of a manipulation to one's self-image. However, we suspect that having all conversation participants use the filter may place everyone on the same level and mitigate feelings of embarrassment or shame in the child condition. Generally, while people had varied reactions to the filters, differences in perception were not found to correlate with gender as had occurred in Chapter 4.

Additionally, we observed that people have different personality traits that may impact their reaction to these manipulations. Looking at the results of the IAT, we see that for instance, the child filter made some people associate themselves more with being youthful, and for others, it had the opposite effect. Such different reactions need to be considered when designing experiences using real-time virtual camera filters.

The real-time camera filters rely on tracking facial features. Therefore, they are not always stable. Occlusion by hands or other objects, as well as extreme head angles

can disrupt the tracking of facial features, which can cause the filter to momentarily disappear. This consequently disrupts the intended illusion.

Furthermore, some participants mentioned that they felt they did not look at their camera stream much, either because they prefer to look away when thinking, prefer not to look at themselves, or prefer to look at their conversation partners. This points to a unique challenge facing the use of real-time filters. During a regular video conferencing session, the illusion of embodiment can only be supported by an image on a two-dimensional screen. This parallels the experience of looking at virtual mirror reflections in immersive VR [37]. However, in contrast to standard immersive VR experiences facilitated using head-mounted displays (HMDs), it is not possible to support the illusion at any viewing angle. Instead, the illusion is limited to the size of the display and can be dispelled when a person looks down at their own physical bodies in the real physical space. Nevertheless, the tendency of the results suggests that the orientation period during which they were instructed to deliberately look at their camera stream may have been enough to make a slight shift.

As limitations to our study, our sample size is small (21 participants), the majority of our participants identified as undergraduate students, and many of the participants considered themselves to be creative. As such, it would be helpful to conduct the study with a different sample population, comprising more people who do not consider themselves creative. Additionally, since the VGT relies solely on providing verbs, language fluency may be a confounding factor. However, the majority of participants were native English speakers (18 of 21), and repeated measures were taken per trial. All in all, our preliminary results suggest that filters may have an impact on some people with regards to feelings of embodiment, mood, and creativity. We see this as an indication for the potential for real-time filters to impact cognition, and see this topic as a fruitful direction for further research.

There may also be other factors at play that can influence people's performance. The novelty effect often accompanies exposure to a new technology. However, since

our participants mostly reported being familiar with real-time virtual camera filters, we believe that this was not a major factor in this study. Another factor may be people's personal preferences. Whether someone finds a visual effect more or less appealing may contribute to their performance during the activity. Further research is needed to understand its level of influence.

5.6 Summarizing Remarks

We performed a novel investigation into the potential impacts of changing one's digital self-image with the use of real-time camera filters in a video-calling scenario. In our user-study with 21 participants, we compared the effects between two different filters, an inventor filter (I) and a child filter (C) against a no-filter condition (A). Participants applied the different videos in a Zoom call and were asked to complete different creativity tasks as well as different surveys. The results of our preliminary study suggest that for some people, real-time virtual camera filters may trigger varying degrees of different moods, levels of embodiment, and creativity. Consequently, we believe that research into the constructive and productive use of real-time camera filters warrants further investigation. In light of the surge in usage of video-based online communication tools due to the COVID-19 pandemic, we believe that the self-view in video calling platforms may serve as a useful channel for stimulation and improvements in cognition.

As we learned that some people have the tendency to avoid looking at their camera image frequently, it would be interesting to explore whether the use of real-time filters by all members in a collaborative video-calling session would amplify any effects. Additionally, we would like to increase the number of participants, study the impact of different filter parameters (e.g. realism), and examine how this approach compares to VR. Another interesting idea would be to consider the possibilities of real-time filters as assistive technology. Filters could be designed to counter personal limitations, such

as an inability to maintain eye-contact due to autism, or mask involuntary tremors due to Parkinson's. We see this wide array of potential applications and extensions to this work as an opportunity worthy of further investigation. Deeper exploration into the use of real-time camera filters is particularly interesting since this approach is highly customizable and scalable.

Chapter 6

Conclusion & Future Work

> "If you really want to change, start with your mindset, attitude, intentions, and how you speak to yourself. Real change starts on the inside."
> - Idil Ahmed

This thesis aimed to explore the possibility of manipulating one's digital self-image as a driver for self-improvement. We conducted an online survey during the COVID-19 pandemic in June and July of 2020 on the use of video-based online communication tools. From the survey, it was found that there is a body of people who are receptive to feedback for self-improvement in this context. The survey furthermore revealed that a visible live video stream of oneself or "*self-view*", is a common and accessible feature in many online communication tools that has the potential to serve as a channel for input and stimulation to users.

Building upon the insights from this survey, we performed two user studies to investigate the potential impact of synthesizing and manipulating one's digital self-image. We explored this concept in the contexts of communication and creativity. In one user study, we leveraged state-of-the-art synthetic media techniques to generate a video of oneself excelling in confident speaking (a target activity), and showed it to people prior to an impromptu speaking task. In the second study, we explored

manipulating one's self-image in real-time to look more like an inventor or more like a child with the use of virtual camera filters while engaging in creativity tasks in an online video call. We collected feedback from participants in these studies to gain insight into the potential effects of manipulating one's digital self-image.

Based on feedback from participants in these studies, we found that exposure to these digitally synthesized and manipulated self-views can unlock different self-beliefs, perceptions, and emotions and that the effects can be quite polarizing. Some felt negatively towards these representations citing reasons such as disliking looking at oneself, finding the representations uncanny, or subjectively disliking the look of these effects. For these participants, they felt such representations can raise insecurities, be distracting, and consequently have detrimental effects on performance for assigned tasks. However, other participants saw them in a positive light and found them to be beneficial. For these participants, synthesized videos of oneself can serve as positive role-models to some, revealing areas for improvement, or giving a sense that they are capable of attaining a greater level of performance. Real-time filters were helpful to some people to reduce feelings of inhibition and self-consciousness. While we studied these effects on a small sample of participants, the preliminary results from these early explorations show the potential for this approach to have positive impacts. We are therefore encouraged to pursue further research in this direction.

It appears that currently, a trade-off exists between the accessibility of these technologies, and the level of immersion that can be supported. While real-time camera filters are already pervasive in the landscape of online communication tools and synthetic media techniques are rapidly becoming more known and accessible, these operate in the realm of 2D visual illusions. This is in contrast to the types of 3D immersive illusions that are facilitated by VR technology. People can easily dispel any illusion by opting to look away from their screen at their real bodies and surroundings. This limits the impact this approach can have, particularly for people who are reluctant to look at digital portrayals of themselves. However, as 3D scanning

and avatar technologies advance, we may see a convergence of this technique and VR where realistic portrayals of ourselves can be used inside immersive VR experiences.

As future work, we would like to explore the impact of personalized synthesized media in different domains. While we focused on confident speech, videos can be generated to reveal and explore many *alternate realities*. For example, a person can be shown a video of oneself being humorous, which could cause a shift in perspective of one's own personality traits. Another application may be to generate videos on-the-fly of a person playing different occupational roles that they otherwise would not be able to imagine themselves in. This may be inspiring for young adults contemplating what they might like to do in the future.

Synthetic media could also be used for interactive and personalized story-telling. For example, videos can be generated to power an immersive choose-your-own-adventure entertainment experience. In a similar vein, real-time camera filters can also be used to present an alternate reality. If people experience a performance boost by employing the use of filters, crafting an experience that highlights this change in behavior or ability can offer a transformative experience that awakens people to the idea that they have had such abilities in them all along. In a similar vein, real-time camera filters may also be able to complement story-telling or story-playing. For instance, in online acting and improvisation classes, people may be able to apply different filters to get inspiration for new ideas and behaviors.

On a grander level, we envision the possibility for *AI-enabled aspirational interfaces* to show people what is possible, and motivate them towards their goals. Mental imagery is a technique often used by high-performance athletes that involves visualizing oneself achieving a desired goal. As developments in artificial intelligence and machine learning have started to enable computers to "dream," we believe that synthetic media can play a role as a tool to assist people to dream big by generating vivid visions to pursue. This concept takes us one step closer towards an ultimate vision of being able to download new abilities in the future.

References

[1] jsPsych. https://www.jspsych.org/. Accessed: 2020-09-17.

[2] Microsoft launches a deepfake detector tool ahead of US election. https://techcrunch.com/2020/09/02/microsoft-launches-a-deepfake-detector-tool-ahead-of-us-election/. Accessed: 2020-09-16.

[3] Teresa M Amabile. The social psychology of creativity: A componential conceptualization. *Journal of personality and social psychology*, 45(2):357, 1983.

[4] Sercan Arik, Jitong Chen, Kainan Peng, Wei Ping, and Yanqi Zhou. Neural voice cloning with a few samples. *Advances in Neural Information Processing Systems*, 31:10019–10029, 2018.

[5] A Nixon deepfake, a 'Moon Disaster' speech and an information ecosystem at risk. https://www.scientificamerican.com/article/a-nixon-deepfake-a-moon-disaster-speech-and-an-information-ecosystem-at-risk1/. Accessed: 2020-09-16.

[6] Kelly J Ashford and Robin C Jackson. Priming as a means of preventing skill failure under pressure. *Journal of Sport and Exercise Psychology*, 32(4):518–536, 2010.

[7] Alan Baddeley, Michael Eysenck, and Michael Anderson. *Memory*. Psychology Press, 2009.

[8] Domna Banakou, Sameer Kishore, and Mel Slater. Virtually being Einstein results in an improvement in cognitive task performance and a decrease in age bias. *Frontiers in psychology*, 9:917, 2018.

[9] Roger E Beaty and Dan Richard Johnson. Automating creativity assessment with SemDis: An open platform for computing semantic distance. *PsyArXiv*, 2020.

[10] E. Bekele, Z. Zheng, A. Swanson, J. Crittendon, Z. Warren, and N. Sarkar. Understanding how adolescents with autism respond to facial expressions in

virtual reality environments. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):711–720, 2013.

[11] Human-computer interaction (HCI). `https://www2.eecs.berkeley.edu/Research/Areas/HCI/`. (Accessed: 14.01.2020).

[12] Philippe Bertrand, Jérôme Guegan, Léonore Robieux, Cade Andrew McCall, and Franck Zenasni. Learning empathy through virtual reality: multiple strategies for training empathy-related abilities using body ownership illusions in embodied virtual reality. *Frontiers in Robotics and AI*, 5:26, 2018.

[13] Paul Boersma and David Weenink. Praat: doing phonetics by computer. `https://www.fon.hum.uva.nl/praat/`. Accessed: 2020-01-18.

[14] Matthew Botvinick and Jonathan Cohen. Rubber hands 'feel' touch that eyes see. *Nature*, 391(6669):756–756, 1998.

[15] LouAnne E Boyd, Saumya Gupta, Sagar B Vikmani, Carlos M Gutierrez, Junxiang Yang, Erik Linstead, and Gillian R Hayes. vrsocial: Toward immersive therapeutic vr systems for children with autism. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[16] LouAnne E Boyd, Alejandro Rangel, Helen Tomimbang, Andrea Conejo-Toledo, Kanika Patel, Monica Tentori, and Gillian R Hayes. Saywat: Augmenting face-to-face conversations for adults with autism. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4872–4883, 2016.

[17] Stéphanie Buisine, Jérôme Guegan, and Frédéric Vernier. Technological innovation in group creativity. In *Creativity, Design Thinking and Interdisciplinarity*, pages 185–201. Springer, 2017.

[18] Stuart K Card. *The psychology of human-computer interaction*. Crc Press, 2018.

[19] Thomas P Carpenter, Ruth Pogacar, Chris Pullig, Michal Kouril, Stephen Aguilar, Jordan LaBouff, Naomi Isenberg, and Alek Chakroff. Survey-software implicit association tests: A methodological and empirical analysis. *Behavior research methods*, 51(5):2194–2208, 2019.

[20] Felix Chang, Mufan Luo, Gregory Walton, Lauren Aguilar, and Jeremy Bailenson. Stereotype threat in virtual learning environments: Effects of avatar gender and sexist behavior on women's math learning outcomes. *Cyberpsychology, Behavior, and Social Networking*, 22(10):634–640, 2019.

[21] Why are younger people more creative than adults? https://slate.com/human-interest/2016/08/why-are-younger-people-more-creative-than-adults.html. Accessed: 2020-01-18.

[22] Kyung Yun Choi and Hiroshi Ishii. ambienBeat: Wrist-worn mobile tactile biofeedback for heart rate rhythmic regulation. In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 17–30, 2020.

[23] Pietro Cipresso, Irene Alice Chicchi Giglioli, Mariano Alcañiz Raya, and Giuseppe Riva. The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Frontiers in psychology*, 9:2086, 2018.

[24] Deepfake Salvador Dalí takes selfies with museum visitors. https://www.theverge.com/2019/5/10/18540953/salvador-dali-lives-deepfake-museum. Accessed: 2020-09-16.

[25] Malaria must die so millions can live. https://www.malariamustdie.com/. Accessed: 2020-09-16.

[26] Project Revoice. https://www.projectrevoice.org/. Accessed: 2020-09-16.

[27] Jean Costa, Malte F Jung, Mary Czerwinski, François Guimbretière, Trinh Le, and Tanzeem Choudhury. Regulating feelings during interpersonal conflicts by changing voice self-perception. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[28] Ionut Damian, Chiew Seng Tan, Tobias Baur, Johannes Schöning, Kris Luyten, and Elisabeth André. Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques. In *Proceedings of the 33rd annual ACM conference on Human factors in computing systems*, pages 565–574, 2015.

[29] Alwin de Rooij, Philip J Corr, and Sara Jones. Creativity and emotion: Enhancing creative thinking by the manipulation of computational feedback to determine emotional intensity. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, pages 148–157, 2017.

[30] Ap Dijksterhuis. Why we are social animals: The high road to imitation as social glue. *Perspectives on imitation: From neuroscience to social science*, 2:207–220, 2005.

[31] M Brent Donnellan, Frederick L Oswald, Brendan M Baird, and Richard E Lucas. The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality. *Psychological assessment*, 18(2):192, 2006.

[32] H Henrik Ehrsson, Nicholas P Holmes, and Richard E Passingham. Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas. *Journal of neuroscience*, 25(45):10564–10573, 2005.

[33] Geraldine Fauville, Jeremy N Bailenson, and Anna Carolina Muller Queiroz. Virtual reality as a promising tool to promote climate change awareness. *Technology and Health*, pages 91–108, 2020.

[34] Here are the companies leading the work-from-home revolution. `https://www.forbes.com/sites/jackkelly/2020/05/24/the-work-from-home-revolution-is-quickly-gaining-momentum/?sh=12aafef81848`. Accessed: 2020-01-14.

[35] Jonas Frich, Michael Mose Biskjaer, and Peter Dalsgaard. Why HCI and creativity research must collaborate to develop new creativity support tools. In *Proceedings of the Technology, Mind, and Society*, pages 1–6. 2018.

[36] Mar Gonzalez-Franco and Tabitha C Peck. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5:74, 2018.

[37] Mar Gonzalez-Franco, Daniel Perez-Marcos, Bernhard Spanlang, and Mel Slater. The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment. In *2010 IEEE virtual reality conference (VR)*, pages 111–114. IEEE, 2010.

[38] Cristina Gonzalez-Liencres, Luis E Zapata, Guillermo Iruretagoyena, Sofia Seinfeld, Lorena Perez-Mendez, Jorge Arroyo-Palacios, David Borland, Mel Slater, and Maria V Sanchez-Vives. Being the victim of intimate partner violence in virtual reality: first-versus third-person perspective. *Frontiers in psychology*, 11:820, 2020.

[39] Anthony G Greenwald, Brian A Nosek, and Mahzarin R Banaji. " understanding and using the implicit association test: I. an improved scoring algorithm": Correction to greenwald et al.(2003). 2003.

[40] Jérôme Guegan, Stéphanie Buisine, Fabrice Mantelet, Nicolas Maranzana, and Frédéric Segonds. Avatar-mediated creativity: When embodying inventors makes engineers more creative. *Computers in Human Behavior*, 61:165–175, 2016.

[41] Jérôme Guegan, Todd Lubart, and Julie Collange. (social) identity and creativity in virtual settings: Review of processes and research agenda. In *The Palgrave Handbook of Social Creativity Research*, pages 191–207. Springer, 2019.

[42] Joy Paul Guilford. The nature of human intelligence. 1967.

[43] Taku Hachisu, Yadong Pan, Soichiro Matsuda, Baptiste Bourreau, and Kenji Suzuki. Facelooks: A smart headband for signaling face-to-face behavior. *Sensors*, 18(7):2066, 2018.

[44] Catherine Hamilton-Giachritsis, Domna Banakou, Manuela Garcia Quiroga, Christos Giachritsis, and Mel Slater. Reducing risk and improving maternal perspective-taking and empathy using virtual embodiment. *Scientific reports*, 8(1):1–10, 2018.

[45] Hal E Hershfield, Daniel G Goldstein, William F Sharpe, Jesse Fox, Leo Yeykelis, Laura L Carstensen, and Jeremy N Bailenson. Increasing saving behavior through age-progressed renderings of the future self. *Journal of Marketing Research*, 48(SPL):S23–S37, 2011.

[46] Mohammed Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W Picard. Mach: My automated conversation coach. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 697–706, 2013.

[47] Adam Haar Horowitz, Tony J Cunningham, Pattie Maes, and Robert Stickgold. Dormio: A targeted dream incubation device. *Consciousness and Cognition*, 83:102938, 2020.

[48] Marco Iacoboni. Imitation, empathy, and mirror neurons. *Annual review of psychology*, 60:653–670, 2009.

[49] Instagram. `https://www.instagram.com/`. Accessed: 2020-01-11.

[50] Saskia Jaarsveld and Thomas Lachmann. Intelligence and creativity in problem solving: the importance of test features in cognition research. *Frontiers in psychology*, 8:134, 2017.

[51] Chris Janiszewski and Robert S Wyer Jr. Content and process priming: A review. *Journal of consumer psychology*, 24(1):96–118, 2014.

[52] Ye Jia, Yu Zhang, Ron Weiss, Quan Wang, Jonathan Shen, Fei Ren, zhifeng Chen, Patrick Nguyen, Ruoming Pang, Ignacio Lopez Moreno, and Yonghui Wu. Transfer learning from speaker verification to multispeaker text-to-speech

synthesis. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31, pages 4480–4490. Curran Associates, Inc., 2018.

[53] Xiaoming Jiang and Marc D Pell. On how the brain decodes vocal cues about speaker confidence. *Cortex*, 66:9–34, 2015.

[54] Xiaoming Jiang and Marc D Pell. The sound of confidence and doubt. *Speech Communication*, 88:106–126, 2017.

[55] Melissa K Jungers and Julie M Hupp. Speech priming: Evidence for rate persistence in unscripted speech. *Language and Cognitive Processes*, 24(4):611–624, 2009.

[56] Jahanvash Karim, Robert Weisz, and Shafiq Ur Rehman. International positive and negative affect schedule short-form (i-panas-sf): Testing for factorial invariance across cultures. *Procedia-Social and Behavioral Sciences*, 15:2016–2022, 2011.

[57] Konstantina Kilteni, Ilias Bergstrom, and Mel Slater. Drumming in immersive virtual reality: the body shapes the way we play. *IEEE transactions on visualization and computer graphics*, 19(4):597–605, 2013.

[58] Kyung Hee Kim. Can we trust creativity tests? a review of the torrance tests of creative thinking (ttct). *Creativity research journal*, 18(1):3–14, 2006.

[59] Charles E Kimble and Steven D Seidel. Vocal signs of confidence. *Journal of Nonverbal Behavior*, 15(2):99–105, 1991.

[60] Martin Kocur, Melanie Kloss, Valentin Schwind, Christian Wolff, and Niels Henze. Flexing muscles in virtual reality: Effects of avatars' muscular appearance on physical performance. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 193–205, 2020.

[61] Tomoko Koda and Pattie Maes. Agents with faces: The effect of personification. In *Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA*, pages 189–194. IEEE, 1996.

[62] Marion Koelle, Abdallah El Ali, Vanessa Cobus, Wilko Heuten, and Susanne CJ Boll. All about acceptability? identifying factors for the adoption of data glasses. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 295–300, 2017.

[63] Marion Koelle, Matthias Kranz, and Andreas Möller. Don't look at me that way! Understanding user attitudes towards data glasses usage. In *Proceedings of the 17th international conference on human-computer interaction with mobile devices and services*, pages 362–372, 2015.

[64] Ellen J Langer. *Counterclockwise: Mindful health and the power of possibility*. Ballantine Books, 2009.

[65] Joanne Leong, Yuehan Wang, Romy Sayah, Stella Rossikopoulou Pappa, Florian Perteneder, and Hiroshi Ishii. Sociabowl: A dynamic table centerpiece to mediate group conversations. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2019.

[66] Sheena Lewis, Mira Dontcheva, and Elizabeth Gerber. Affective computational priming and creativity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 735–744, 2011.

[67] Manon Marinussen and Alwin de Rooij. Being yourself to be creative: How self-similar avatars can support the generation of original ideas in virtual environments. In *Proceedings of the 2019 on Creativity and Cognition*, pages 285–293. 2019.

[68] David M Markowitz, Rob Laha, Brian P Perone, Roy D Pea, and Jeremy N Bailenson. Immersive virtual reality field trips facilitate learning about climate change. *Frontiers in psychology*, 9:2364, 2018.

[69] Ilaria Minio-Paluello, Giuseppina Porciello, Marco Gandolfo, Sarah Boukarras, and Salvatore M Aglioti. The enfacement illusion boosts facial mimicry. *Cortex*, 123:113–123, 2020.

[70] Yisroel Mirsky and Wenke Lee. The creation and detection of deepfakes: A survey. *arXiv preprint arXiv:2004.11138*, 2020.

[71] Masahiro Mori, Karl F MacDorman, and Norri Kageki. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100, 2012.

[72] Yondu Mori and Marc D Pell. The look of (un) confidence: Visual markers for inferring speaker confidence in speech. *Frontiers in Communication*, 4:63, 2019.

[73] Naoto Nakazato, Shigeo Yoshida, Sho Sakurai, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. Smart face: enhancing creativity during video conferences using real-time facial deformation. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 75–83, 2014.

[74] Solène Neyret, Xavi Navarro, Alejandro Beacco, Ramon Oliva, Pierre Bourdin, Jose Valenzuela, Itxaso Barberia, and Mel Slater. An embodied perspective as a victim of sexual harassment in virtual reality reduces action conformity in a later milgram obedience scenario. *Scientific reports*, 10(1):1–18, 2020.

[75] Thanh Thi Nguyen, Cuong M Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, and Saeid Nahavandi. Deep learning for deepfakes creation and detection. *arXiv preprint arXiv:1909.11573*, 2019.

[76] Jonas Oppenlaender and Simo Hosio. Design recommendations for augmenting creative tasks with computational priming. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, pages 1–13, 2019.

[77] Orai. `https://www.orai.com/`. Accessed: 2020-09-09.

[78] Francesco Pagnini, Cesare Cavalera, Eleonora Volpato, Benedetta Comazzi, Francesco Vailati Riboni, Chiara Valota, Katherine Bercovitz, Enrico Molinari, Paolo Banfi, Deborah Phillips, et al. Ageing as a mindset: a study protocol to rejuvenate older adults with a counterclockwise psychological intervention. *BMJ open*, 9(7):e030411, 2019.

[79] Konstantin A Pantserev. The malicious use of AI-based deepfake technology as the new threat to psychological security and political stability. In *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity*, pages 37–55. Springer, 2020.

[80] Tabitha C Peck, Jessica J Good, and Kimberly A Bourne. Inducing and mitigating stereotype threat through gendered virtual body-swap illusions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[81] Tabitha C Peck, Jessica J Good, and Kimberly A Bourne. Inducing and mitigating stereotype threat through gendered virtual body-swap illusions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[82] Say hello to presenter coach, powerpoint's new ai-powered tool which will help you nail your next presentation. `https://news.microsoft.com/europe/2019/06/18/say-hello-to-presenter-coach-powerpoints-new-ai-powered-tool-which-will-help-you-nail-your-next-presentation/`. Accessed: 2020-09-09.

[83] Ranjani Prabhakaran, Adam E Green, and Jeremy R Gray. Thin slices of creativity: Using single-word utterances to assess creative cognition. *Behavior research methods*, 46(3):641–659, 2014.

[84] Presentr. "https://presentr.me/". Accessed: 2020-09-09.

[85] Qualtrics. https://www.qualtrics.com/. Accessed: 2020-01-11.

[86] Mark A Runco and Garrett J Jaeger. The standard definition of creativity. *Creativity research journal*, 24(1):92–96, 2012.

[87] Samiha Samrose, Ru Zhao, Jeffery White, Vivian Li, Luis Nova, Yichen Lu, Mohammad Rafayet Ali, and Mohammed Ehsan Hoque. Coco: Collaboration coach for understanding team dynamics during video conferencing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–24, 2018.

[88] Aliaksandr Siarohin, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci, and Nicu Sebe. First order motion model for image animation. In *Advances in Neural Information Processing Systems*, pages 7137–7147, 2019.

[89] Mel Slater, David-Paul Pertaub, Chris Barker, and David M Clark. An experimental study on fear of public speaking using a virtual environment. *CyberPsychology & Behavior*, 9(5):627–633, 2006.

[90] Snap camera. https://snapcamera.snapchat.com/. Accessed: 2020-01-11.

[91] Snapchat. https://www.snapchat.com/. Accessed: 2020-01-11.

[92] Spatial. https://spatial.io/. Accessed: 2020-01-14.

[93] Speeko. https://www.speeko.co/. Accessed: 2020-09-09.

[94] Barry E Stein and Terrence R Stanford. Multisensory integration: current issues from the perspective of the single neuron. *Nature reviews neuroscience*, 9(4):255–266, 2008.

[95] J Ridley Stroop. Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6):643, 1935.

[96] Supasorn Suwajanakorn, Steven M Seitz, and Ira Kemelmacher-Shlizerman. Synthesizing obama: learning lip sync from audio. *ACM Transactions on Graphics (TOG)*, 36(4):1–13, 2017.

[97] Synthetic media. https://en.wikipedia.org/wiki/Synthetic_media. Accessed: 2020-01-18.

[98] Ana Tajadura-Jiménez, Domna Banakou, Nadia Bianchi-Berthouze, and Mel Slater. Embodiment in a child-like talking virtual body influences object size perception, self-identification, and subsequent real speaking. *Scientific Reports*, 7(1):1–12, 2017.

[99] M Iftekhar Tanveer, Emy Lin, and Mohammed Hoque. Rhema: A real-time in-situ intelligent interface to help people with public speaking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 286–295, 2015.

[100] Elizabeth R Tenney, Barbara A Spellman, and Robert J MacCoun. The benefits of knowing what you know (and what you don't): How calibration affects credibility. *Journal of Experimental Social Psychology*, 44(5):1368–1375, 2008.

[101] Kristen M Tooley, Agnieszka E Konopka, and Duane G Watson. Can intonational phrase structure be primed (like syntactic structure)? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2):348, 2014.

[102] Kristen M Tooley, Agnieszka E Konopka, and Duane G Watson. Assessing priming for prosodic representations: Speaking rate, intonational phrase boundaries, and pitch accenting. *Memory & cognition*, 46(4):625–641, 2018.

[103] Manos Tsakiris. Looking for myself: current multisensory input alters self-face recognition. *PloS one*, 3(12):e4040, 2008.

[104] Rafael Valle, Jason Li, Ryan Prenger, and Bryan Catanzaro. Mellotron: Multispeaker expressive voice synthesis by conditioning on rhythm, pitch and global style tokens. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6189–6193. IEEE, 2020.

[105] Oshin Vartanian, Ingrid Smith, Timothy K Lam, Kristen King, Quan Lam, and Erin L Beatty. The relationship between methods of scoring the alternate uses task and the neural correlates of divergent thinking: Evidence from voxel-based morphometry. *NeuroImage*, 223:117325, 2020.

[106] Jinping Wang, Hyun Yang, Ruosi Shao, Saeed Abdullah, and S Shyam Sundar. Alexa as coach: Leveraging smart speakers to build social agents that reduce public speaking anxiety. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[107] Brenda K Wiederhold. Connecting through technology during the coronavirus disease 2019 pandemic: Avoiding "Zoom Fatigue", 2020.

[108] Rebecca A Williamson, Vikram K Jaswal, and Andrew N Meltzoff. Learning the rules: Observation and imitation of a sorting strategy by 36-month-old children. *Developmental psychology*, 46(1):57, 2010.

[109] Nick Yee and Jeremy Bailenson. The Proteus effect: The effect of transformed self-representation on behavior. *Human communication research*, 33(3):271–290, 2007.

[110] Nick Yee, Jeremy N Bailenson, and Nicolas Ducheneaut. The proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Communication Research*, 36(2):285–312, 2009.

[111] Jayden Ziegler, Giulia Bencini, Adele Goldberg, and Jesse Snedeker. How abstract is syntax? Evidence from structural priming. *Cognition*, 193:104045, 2019.

[112] Zoom. `https://zoom.us/`. Accessed: 2020-01-11.

[113] Filters, reactions, lighting more! New features to liven up your meetings. `https://blog.zoom.us/filters-reactions-lighting-features-zoom-meetings-2/`. Accessed: 2020-01-11.

Appendix A

Online Communication Survey

The original survey was distributed online via Google Forms. The version here is the native pdf export of the survey.

Survey on Online Communication

We would like to understand your experience with online video-based communication platforms for professional activities.

* Required

Consent to participate

The purpose of this research is to understand people's experiences with online video calling for professional activities. Results from this survey will be be used as part of an MIT Master's thesis and may be used in future academic research publications.

Please note that:
1. Your participation is voluntary.
2. Some questions are mandatory. However, if you do not want to answer them, you can exit the survey.
3. At any time, you may decline further participation without adverse consequences. To do this, simply close out of the survey without submitting your answers.
4. Your confidentiality and/or anonymity are assured. We will not collect any personally identifying information (e.g. your name or contact information)

If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143B, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253 6787.

1. Do you consent to participate? *

Check all that apply.

- [ ] I understand and consent to the terms of participation for this survey.

Online communication

2. Do you use **video-calling** for your work, studies, or other professional activities? *

Mark only one oval.

- ◯ Yes
- ◯ No

Using video-calling platforms for work (1/4)

3. For **WORK** purposes, which **VIDEO** calling platforms do you use? *

Check all that apply.

- [ ] Zoom
- [ ] Skype
- [ ] Google Hangouts
- [ ] Facebook Messenger
- [ ] WhatsApp
- [ ] Microsoft Teams
- [ ] BlueJeans by Verizon

Other: [ ] _____

4. For **WORK** purposes, which device(s) do you use for holding **VIDEO** calls? *

Check all that apply.

- [ ] Mobile
- [ ] Tablet
- [ ] Laptop / Desktop

5. How often do you use a video-based communication platform? *

Mark only one oval.

- ◯ Less than once a week
- ◯ A few times each week
- ◯ Every day
- ◯ Multiple times daily
- ◯ Continuously

6. During video calls, do you watch the other conversation participants? *

Mark only one oval.

	1	2	3	4	5	
I never look at them.	◯	◯	◯	◯	◯	I'm almost always looking at others videos.

7. Do you look at the video that shows yourself? (aka. "Self-View") *

Mark only one oval.

- ◯ I never do
- ◯ Only at the beginning
- ◯ Once in a while
- ◯ Quite frequently
- ◯ Almost constantly

8. If you check your self-view, why? Check all that apply. Add your own response if it is not listed.

Check all that apply.

- ☐ Check if I'm in the frame.
- ☐ Check that my background/lighting is appropriate.
- ☐ Check that my physical appearance is appropriate.
- ☐ Check that my facial expression and gestures are appropriate.
- ☐ For no particular reason

Other: ☐ _____

9. How frequently do you turn off your video stream, such that no one can see you while you talk? *

Mark only one oval.

	1	2	3	4	5	
I never have my camera on.	◯	◯	◯	◯	◯	I always have my camera on.

10. If you at any time turn off your video stream, why?

_____

_____

_____

_____

_____

11. Can you disable your self-view in the communication platform you use? This means you don't see yourself, but others still can. *

Mark only one oval.

- ◯ Yes, I can disable my self-view in one or more platforms I use.
- ◯ No, I cannot disable my self-view in any of the platforms I use.
- ◯ I don't know.

12. If you are able to, how frequently do you turn off your self-view?

Mark only one oval.

	1	2	3	4	5	
I always have my self-view OFF/DISABLED.	◯	◯	◯	◯	◯	I always leave my self-view ON/ENABLED.

13. If you at any time turn off your self-view, why?

_____

_____

_____

_____

_____

14. How often is screen-sharing used during calls you attend? *

Mark only one oval.

	1	2	3	4	5	
Screen sharing is never used.	◯	◯	◯	◯	◯	Screen sharing is always used.

15. If screen-sharing is used, how frequently do you look at your SELF-VIEW?

Mark only one oval.

	1	2	3	4	5	
I never look at my self-view.	◯	◯	◯	◯	◯	I always look at my self-view.

Speaking with confidence in video calls (2/4)

16. Are you satisfied with your communication abilities in video calls? *

Mark only one oval.

	1	2	3	4	5	
Not at all satisfied.	◯	◯	◯	◯	◯	Completely satisfied.

17. If you're not completely satisfied, what do you wish you could improve on? If you are, write "N/A". *

18. To what degree do you believe it's important to speak confidently on your calls? *

Mark only one oval.

	1	2	3	4	5	
Not at all.	◯	◯	◯	◯	◯	Very important.

19. What do you do to look and sound confident?

20. What is more difficult in video calls compared to in-person conversations?

21. What is easier in video calls compared to in-person conversations?

22. Would you like suggestions for improving your speaking skills in video calls? *

Mark only one oval.

	1	2	3	4	5	
No, not at all.	◯	◯	◯	◯	◯	Yes, absolutely.

23. What types of information about yourself are you interested in, to help you speak clearly and confidently? Check all that apply. *

Check all that apply.

- [ ] Posture
- [ ] Physical Gestures
- [ ] Facial Expressions
- [ ] Voice Features (e.g. speed, pitch, pace, etc.)
- [ ] Choice of words
- [ ] Order of words / structure
- [ ] Level of participation in the conversation
- [ ] Other people's responses to me
- [ ] None
- [ ] Other: _____

24. What types of guidance or feedback would you be interested in? Check all that apply. *

Check all that apply.

- [ ] Advice/Tips before a conversation
- [ ] Suggestions during the conversation
- [ ] Summary of your speaking performance after the conversation
- [ ] Audio hints or cues (that only you can hear)
- [ ] Visual hints or cues (that only you can see)
- [ ] Haptic hints or cues (that only you can feel)
- [ ] None
- [ ] Other: _____

Types of calls (3/4)

25. What types of video-calls do you have? Check all that apply.

Check all that apply.

- [ ] Status Update: discuss updates, progress, challenges and/or next steps
- [ ] Information Sharing: share info in presentations, brown bag sessions, panels, lectures
- [ ] Planning / Decision Making: evaluate options, rank preferences, decide course of action
- [ ] Problem Solving / Brainstorming: address an identified problem
- [ ] Team Building: strengthen relationships
- [ ] Practice a Skill: perform activities to improve an ability
- [ ] Other: _____

26. How often do you actively participate (i.e. speak) on calls of different sizes?

Mark only one oval per row.

	Never	Less than once a week	Once a week	A few times a week	Every day	Multiple times a day
1 on 1	◯	◯	◯	◯	◯	◯
Small Groups (<5)	◯	◯	◯	◯	◯	◯
Medium Groups (6-10)	◯	◯	◯	◯	◯	◯
Large Groups (11-25)	◯	◯	◯	◯	◯	◯
Very Large (>25)	◯	◯	◯	◯	◯	◯

27. How frequently do you speak IMPROMPTU in a call? (i.e. you spontaneously say ideas/opinions/feedback, without practicing what to say beforehand) *

Mark only one oval.

- ◯ Never
- ◯ Less than once a week
- ◯ A few times a week
- ◯ Every day
- ◯ Multiple times a day

28. How frequently do you REHEARSE what you will say before a call? (i.e. you practice what you'll say beforehand) *

Mark only one oval.

- ◯ Never
- ◯ Less than once a week
- ◯ A few times a week
- ◯ Every day
- ◯ Multiple times a day

A bit about you (4/4)

29. Age: *

_____

30. Gender:

Check all that apply.

- ☐ Female
- ☐ Male
- ☐ Prefer not to disclose
- Other: ☐ _____

31. Occupation (e.g. graduate student, web developer, teacher): *

_____

Thank you for completing this survey! You can now submit your answers.

Appendix B

Communication Study Pre-Survey

The original survey was distributed online via Google Forms. The version here is the native pdf export of the survey.

Clear and Confident Speech Study: Pre-Survey

Thank you for volunteering to participate in this study! Please complete the questions below. Once custom material has been prepared for you, we will email you a unique link to a website to complete the study.

If there are any issues with the photo you have uploaded, we will also email you to get in contact. If you have questions, please feel free to email joaleong@media.mit.edu.

* Required

1. Email address *

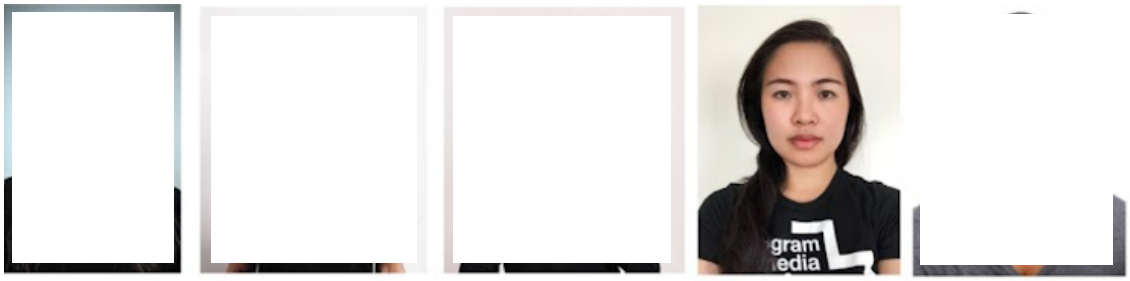
_____

2. Your full name: *

_____

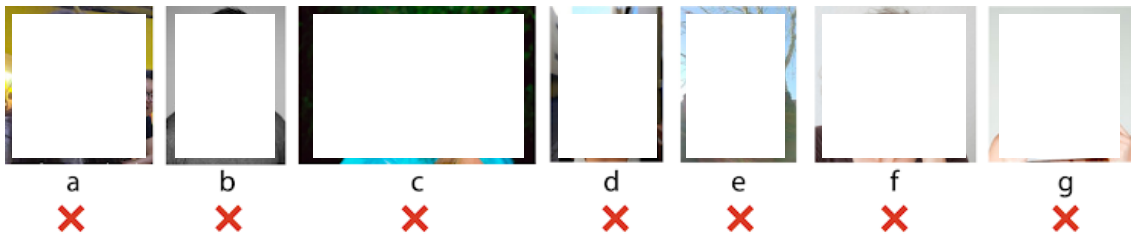
Please upload a photo of yourself...

Submit a color photo of yourself. Please look straight at the camera with your mouth and lips closed (i.e. no smiling!). Your head and the tops of your shoulders should be clearly visible. The photo should have a plain background. No hands or objects should be touching or obscuring the face (however glasses are OK if they do not have strong reflections).



*Some images removed for privacy protection.

These are examples of types of photos that are unsuitable, for the following reasons: A) There is another person and her mouth is open B) it is black and white. B and C) the head is turned / it is taken at an extreme angle. D and E) very poor lighting and busy background, F) a hand is touching her face and she is smiling, G) parts of her face are covered.



*Some images removed for privacy protection.

3. Upload a portrait photo of yourself here: *

Files submitted:

4. Which gender do you most identify as? *

Mark only one oval.

- Male
- Female
- Other / Prefer not to say

5. Age: *

_____

6. Occupation: *

_____

7. How would you rate your oral communication skills? *

Mark only one oval.

	1	2	3	4	5	
Very Weak						Very Strong

8. How confident do you feel in impromptu speaking situations? (E.g. speaking up during a meeting without preparation) *

Mark only one oval.

	1	2	3	4	5	
Very Unconfident	◯	◯	◯	◯	◯	Very Confident

9. Do you think seeing videos of yourself speaking well would boost your self-confidence in your ability to speak? *

Mark only one oval.

	1	2	3	4	5	
Not At All	◯	◯	◯	◯	◯	Definitely

Appendix C

Communication Study Task Instructions

These were instructions and prompts used as part of the assigned tasks in the study "Exploring Synthetic Media for Confident Speech."

Communication Study – Initial Instructions

Imagine you are being interviewed for a job position you would like to get.

You will be asked a series of questions.

- Answer these questions as clearly and confidently as you can.
- Your answers will be recorded on camera.
- Experts will review your videos. You will be assessed on your communication skills based on your responses.

Please note the following:

- You will see a few short video clips as part of the process. This may include videos of artwork, someone speaking, or even a generated video of you speaking. We would like to know if this can be used to positively impact one's speaking habits.
- While your responses will be reviewed and assessed by study experts, the results from this will not impact you in real life.
- Please use a laptop or desktop computer to complete this study.

Thank you for volunteering to participate in this study.

Click the button below to begin! [START]

Communication Study – Interview Questions

1. How do you stay organized if there are many tasks to keep track of? (Practice)
2. What are your preferred tools for online communication, and why? (Factual)
3. What are your responsibilities in your current/most recent job? (Factual)
4. What are some important traits for a teammate to have, and why? (Factual)
5. Who has been an influential person in your career and why? (Open-Ended)
6. What is more important for success - talent or hard work? (Open-Ended)
7. What impact do you think the outcome of the US presidential election will have on your career in the next 4 years? (Open-Ended)

Communication Study – Trial Survey Questions

5-point Likert Scale Questions. Answers ranged from 1 – 5.

- Rate how stressed you felt answering the question. (Stress)
  - Not at all stressed, Extremely Stressed
- Rate how clearly you spoke when answering the question. (Clarity)
  - Very unclearly, Very clearly
- Rate how confidently you spoke when answering the question. (Confidence)
  - Very unconfidently, Very confidently
- Rate how comfortable you were with the question topic. (Comfort)
  - Very uncomfortable, Very comfortable
- Rate how satisfied you were with how you answered the question. (Satisfaction)
  - Very unsatisfied, Very satisfied

5-point Likert Scale Questions. Answers ranged from 1 – 5.

- Rate how much the person in the video looked like you. (Visual Resemblance:)
  - It wasn't at all like me, It was exactly like me
- Rate how much the person in the video sounded like you (Auditory Resemblance)
  - It wasn't at all like me, It was exactly like me
- Rate how similar the speaking style in the video was to yours. (Similarity)
  - Completely different, Exactly the same
- Rate how much you like the speaking style presented in the video (Style)
  - I didn't like it all, I like it very much
- Do you think the video affected your speaking style positively or negatively? (Impact)
  - Very negatively, Very positively
- Rate the visual and audio quality of the video (Video Quality)
  - Poor quality, Excellent quality

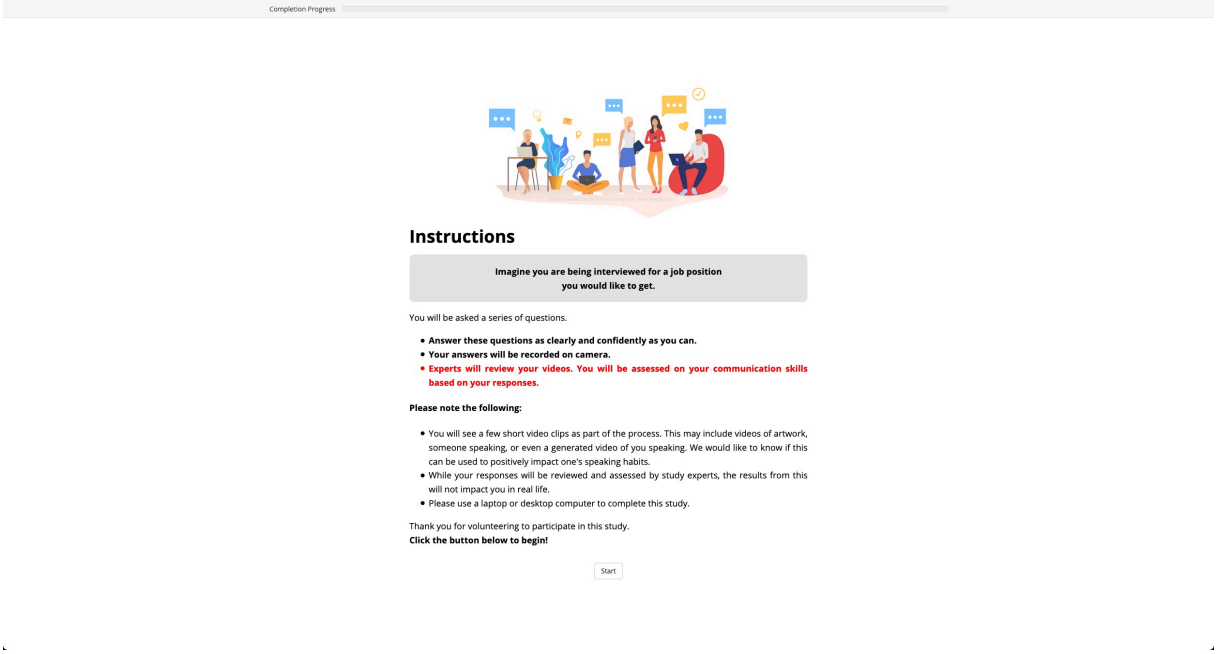
Communication Study – Trial Survey Questions

(Open Questions)

- What did you think of seeing a video of another speaker before you spoke?
- What did you think of seeing a video of yourself before you spoke?
- Do you have any other comments?

Communication Study – Web Interface Sample Screenshots

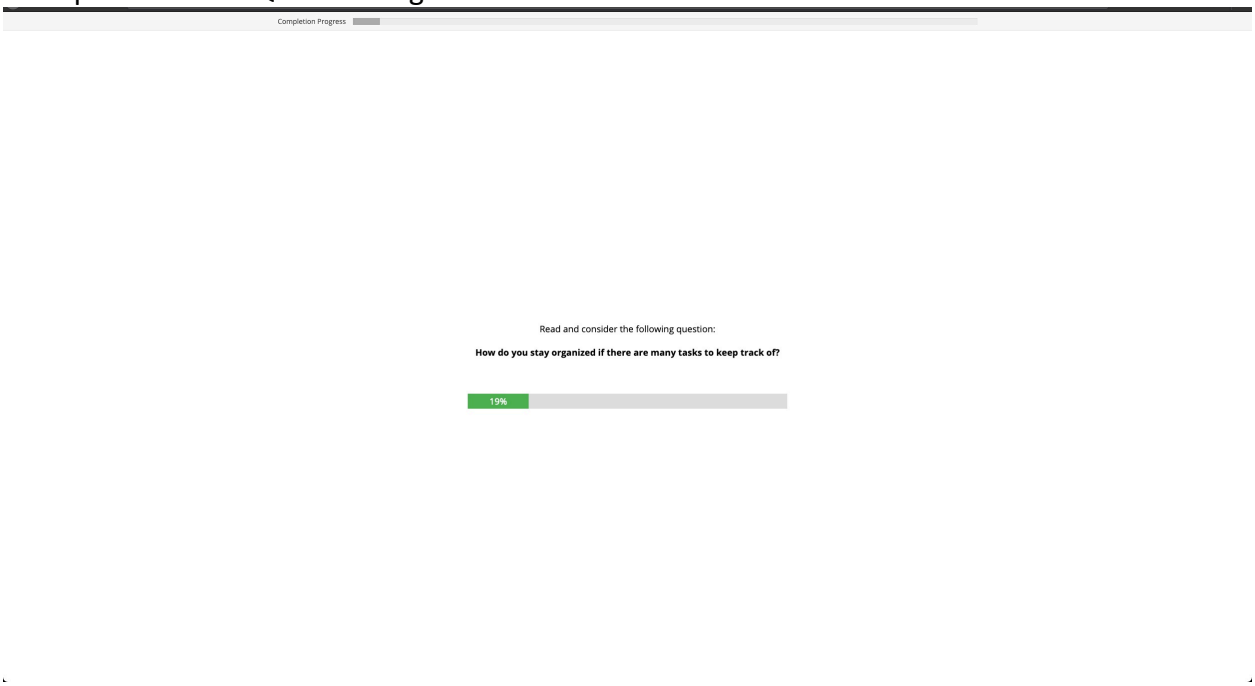
Sample Instruction Page:



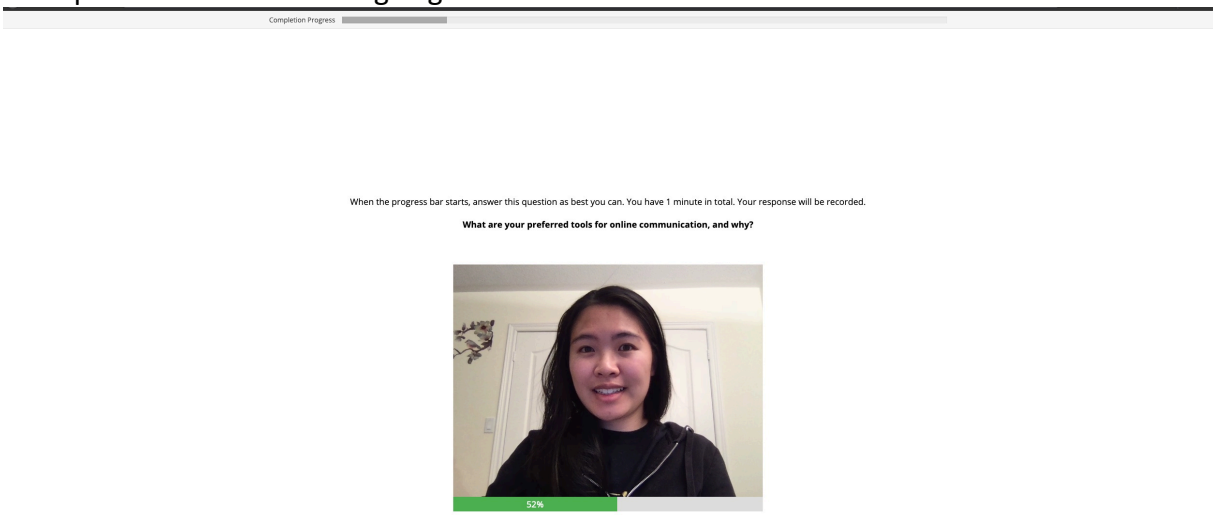
Example “None” Video Page:

Example of a “Natural”/”Generated” Speaker Video Page (Face blocked for privacy reasons):

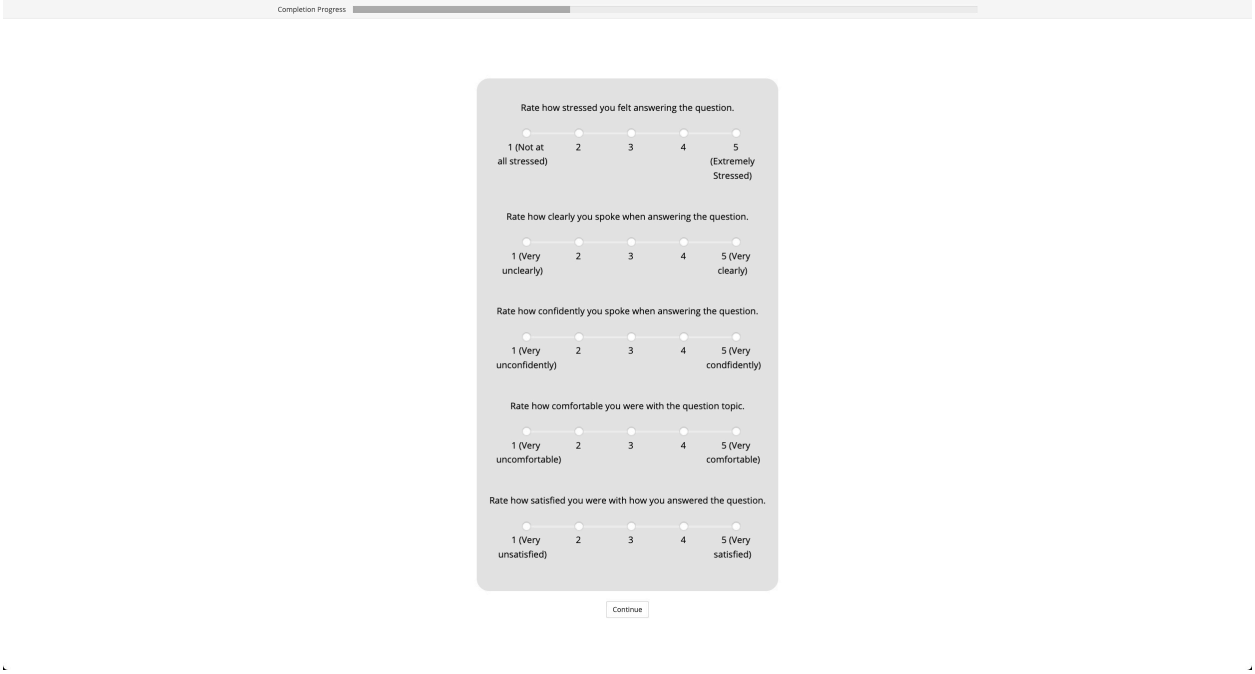
Example Interview Question Page:

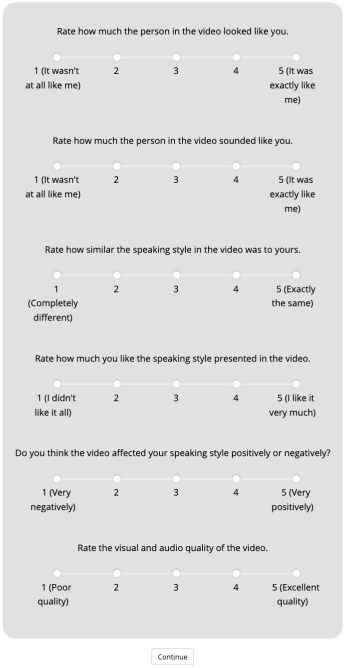


Example Interview Answering Page:

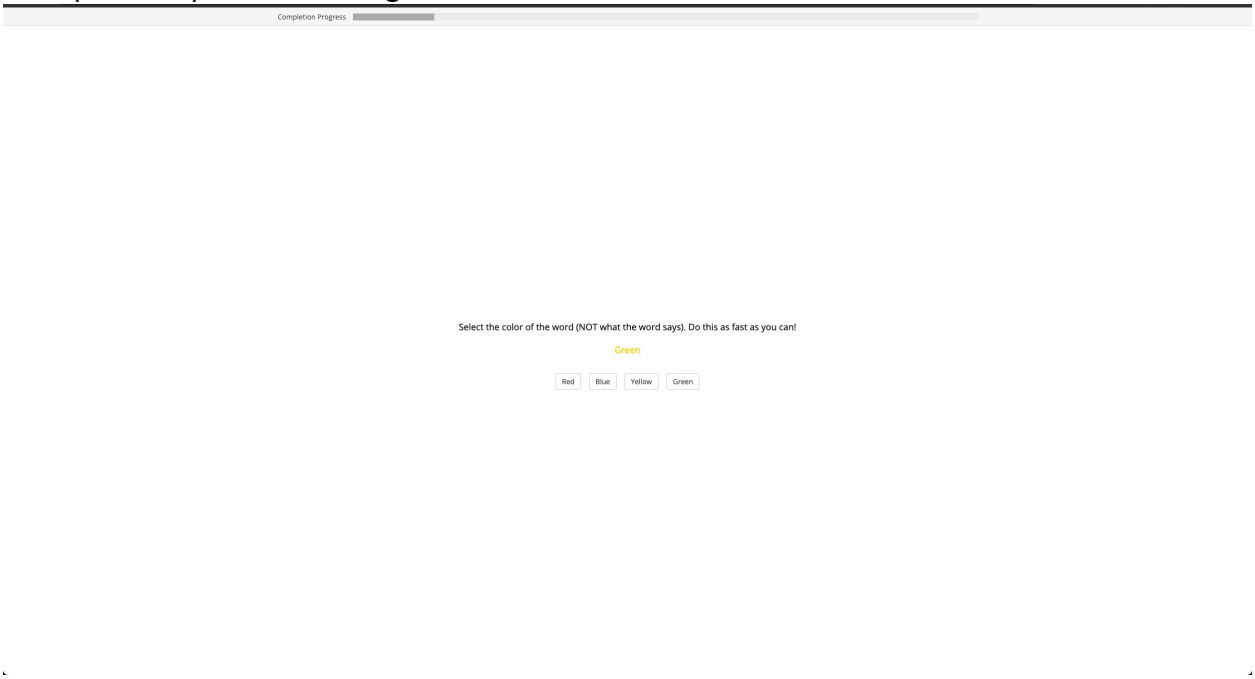


Examples of Survey Pages:





Example Stroop Color Test Page:



Appendix D

Creativity Study Pre-Survey

The original survey was distributed online via Google Forms. The version here is the native pdf export of the survey.

Pre-Survey

Thank you for agreeing to participate in this study! If you have any questions, please contact the study coordinator at joaleong@media.mit.edu.

For each statement, indicate how accurately you believe it describes you, as you are now.

1 = Very Inaccurate
2 = Moderately Inaccurate
3 = Neither Accurate Nor Inaccurate
4 = Moderately Accurate
5 = Very Accurate as a description of you.
* Required

1. 1. I am the life of the party. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

2. 2. I sympathize with others' feelings. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

3. 3. I get chores done right away. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

4. 4. I have frequent mood swings. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

5. 5. I have a vivid imagination. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

6. 6. I don't talk a lot. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

7. 7. I am not interested in other people's problems. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

8. 8. I often forget to put things back in their proper place. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

9. 9. I am relaxed most of the time. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

10. 10. I am not interested in abstract ideas. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

11. 11. I talk to a lot of different people at parties. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

12. 12. I feel others' emotions. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

13. 13. I like order. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

14. 14. I get upset easily. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

15. 15. I have difficulty understanding abstract ideas. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

16. 16. I keep in the background. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

17. 17. I am not really interested in others. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

18. 18. I make a mess of things. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate	◯	◯	◯	◯	◯	Very Accurate

19. 19. I seldom feel blue. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate						Very Accurate

20. 20. I do not have a good imagination. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate						Very Accurate

21. 21. I am creative. *

Mark only one oval.

	1	2	3	4	5	
Very Inaccurate.						Very Accurate.

A bit about you

*** This data is used exclusively to personalize your study experience.

22. Full Name*** *

23. Age *

24. Gender *

Mark only one oval.

- Female
- Male
- Prefer not to say
- Other: _____

25. Occupation (If you are a student, please also specify what you study. E.g. Student, Architecture) *

_____

26. Native Language *

_____

27. Home Town/City***

_____

28. Home Country*** *

_____

29. How experienced are you with using "camera filters" from applications (e.g. from apps such as Snapchat or Instagram)? *

Mark only one oval.

	1	2	3	4	5	
Not Familiar						Very Familiar

Appendix E

Creativity Study Task Instructions

These were instructions and prompts used as part of the assigned tasks in the study "Exploring Real-Time Camera Filters fo Creativity."

Creativity Study – Alternate Uses Task Instructions

Instructions: Think of as many unique and unusual uses for the object I will give you. Please explain your ideas in such a way that someone else listening to your idea would understand exactly what to do with the object. You will have 2 minutes to brainstorm for each object. I will nod when I understand your idea, which will let you know that you can continue. I will wave to let you know when the time has run out.

Objects: *Pizza Box, Water Bottle, Broom, Paper Clip, Plastic Fork, Rubber Band*

Creativity Study – Verb Generation Task Instructions & Word Lists

Instructions: I have a list of nouns. Your mission is to say a single verb that could be associated with each noun. For example, given the word "scissors," you could say "cut," because you can cut something with the scissors. You could also say "throw" because you could technically throw them too.

X LIST	Frequency	Y LIST	Frequency	Z LIST	Frequency
office	0.606	debt	0.606	horn	0.606
soap	0.592	drug	0.592	pillow	0.592
belt	0.535	pill	0.535	sofa	0.535
rock	0.521	oath	0.521	finger	0.521
blade	0.521	shovel	0.521	feet	0.521
note	0.507	soup	0.507	baby	0.507
artist	0.479	tool	0.479	money	0.479
card	0.465	poem	0.465	rose	0.465
leaf	0.437	store	0.437	music	0.437
lamp	0.423	boot	0.423	drum	0.423
street	0.409	canoe	0.408	fist	0.408
paper	0.408	phone	0.408	glass	0.408
taxi	0.38	house	0.38	oven	0.38
tongue	0.38	snow	0.38	café	0.38
muscle	0.38	ring	0.38	pan	0.38
home	0.352	dish	0.352	infant	0.352

*These words were taken from the original set found in Prabhakaran, Ranjani, Adam E. Green, and Jeremy R. Gray. "Thin slices of creativity: Using single-word utterances to assess creative cognition." *Behavior research methods* 46.3 (2014): 641-659.

Creativity Study – Instructions for Orientation Task

Instructions: Please follow my lead by copying me:

1. Wave one hand.
2. Wave the other hand.
3. Shrug your shoulders.
4. Touch (the tops of) your shoulders.
5. Touch the top of your head.
6. Pat one shoulder (cross-body).
7. Pat the other shoulder (cross-body).
8. Nod your head.
9. Shake your head.
10. Smile.
11. Frown.
12. Puff up your cheeks.
13. Raise your eyebrows up and down.
14. Touch your forehead.
15. Touch the tip of your nose.
16. Touch your cheeks.

Continue to move around, similar to how we just did, and experiment with what you see in your camera stream. (20 seconds)

Appendix F

Creativity Study Trial Survey

The original survey was distributed online via Qualtrics. The version here is the native pdf export of the survey.

Subject & Opinions

Please answer each of the following questions. If you need help, please ask the study coordinator.

Participant ID Number

How difficult was it for you to think of ideas for:

	Very Difficult	Somewhat Difficult	Neutral	Somewhat Easy	Very Easy
Object 1	○	○	○	○	○
Object 2	○	○	○	○	○

How much did you like your video stream?

	I did not like it at all.	I somewhat disliked it.	I neither liked it nor disliked it.	I somewhat liked it.	I liked it very much.
My Video Stream	○	○	○	○	○

Mood

Think about how you felt during the past activity. To what extent did you feel:

	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
Upset	○	○	○	○	○
Hostile	○	○	○	○	○
Alert	○	○	○	○	○
Ashamed	○	○	○	○	○
Inspired	○	○	○	○	○
Nervous	○	○	○	○	○
Determined	○	○	○	○	○
Attentive	○	○	○	○	○
Afraid	○	○	○	○	○
Active	○	○	○	○	○

Please rate how you feel:

	Very Sad		Very Happy	
0	25	50	75	100

Embodiment

Body Ownership

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
Q1. I felt as if the person I saw on the screen was myself.	○	○	○	○	○	○	○
Q2. It felt as if the person I saw on the screen was someone else.	○	○	○	○	○	○	○
Q3. It felt as if I might have more than one body.	○	○	○	○	○	○	○
Q4. I felt as if the face I saw when looking at my screen was my own face.	○	○	○	○	○	○	○
Q5. I felt as if the face I saw when looking at my screen was another person.	○	○	○	○	○	○	○

Agency and Motor Control

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
Q6. It felt like I could control the person on the screen as if it were myself.	○	○	○	○	○	○	○
Q7. The movements of the person on my screen were caused by my movements.	○	○	○	○	○	○	○
Q9. I felt as if the person I saw on screen was moving by itself.	○	○	○	○	○	○	○

Tactile Sensations

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
Q10. It seemed as if I felt the touch of my hand in the location where I saw the hand touch me on screen.	O	O	O	O	O	O	O
Q11. It seemed as if the touch I felt was located somewhere between my physical self and the on-screen person.	O	O	O	O	O	O	O

Location of the Body

	Strongly Disagree	Disagree	Somewhat disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
Q14. I felt as if my body was located where I saw the virtual person.	O	O	O	O	O	O	O
Q15. I felt out of my body.	O	O	O	O	O	O	O
Q16. I felt as if my real body was moving with the on-screen body, or as if the on-screen body were moving with my real body.	O	O	O	O	O	O	O

External Appearance

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
Q17. It felt as if my real body was turning into an 'avatar' body.	O	O	O	O	O	O	O
Q18. At some point it felt as if my real self was starting to take on the posture or shape of the person I saw on-screen.	O	O	O	O	O	O	O

	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
Q19. At some point it felt like the on-screen person resembled my real self, in terms of shape, skin tone or other visual features.	○	○	○	○	○	○	○
Q20. It felt like I was wearing a mask or different accessories from when I entered the study.	○	○	○	○	○	○	○

Block 3

Do you think you looked younger, older, or the same age?

- ○ Younger
- ○ Older
- ○ Same Age

Please provide a number for the age you think you looked:

Appendix G

Creativity Study Exit Survey

The original survey was distributed online via Qualtrics. The version here is the native pdf export of the survey.

Default Question Block

Please answer each of the following questions. If you need help, ask the study coordinator.

Participant ID Number

	1 (Not at All)	2	3	4	5	6	7 (Very Much)
Do you consider yourself to be creative?	○	○	○	○	○	○	○

	1 (Children Perform Better)	2	3	4	5	6	7 (Adults Perform Better)
If a creativity test were to show age differences, who do you think would perform better?	○	○	○	○	○	○	○

	1 (No Stereotype)	2	3	4	5	6	7 (Very Strong Stereotype)
To what extent do you think there is a stereotype in the U.S. that children are more creative compared to adults?	○	○	○	○	○	○	○

Rate how much you

	1 (Not at All)	2	3	4	5	6	7 (Very Much)
...are able to 'act' different roles	○	○	○	○	○	○	○
..consider yourself to be playful	○	○	○	○	○	○	○