School of Humanities, Media and Creative Communication
Candidate: Milan Marić - Markuza
Supervisors: Dr Claire Henry and Dr Ian Goodwin

The evolving aesthetics of the moving image in vertical video online

Table of Contents

Glossary

Terms in the thesis that can be found in this glossary are marked with a superscript x letter.

Adobe Premiere – Timeline-based video editing software.

Adobe Premiere optical flow filter – The optical flow filter estimates the motion between two frames of video and renders an intermediate frame that interpolates the motion.

Articulation of the Z-axis – In context of moving images, Z-axis is an imaginary line that extends from the front of the camera lens to the horizon. Skilful articulation of Z-axis increase the impression of three dimensional space in two-dimensional medium such as video.

Attentional synchrony – Attentional synchrony is a theory in the studies of scene perception, proposing that multiple viewers of dynamic scenes attend to the same location of the scene at the same time.

Audio meter – Part of the application or audio/video recording hardware that measures audio input levels in dB value.

Audio wild-track – A recording of dialogue or sound on the filming set but without the camera running. The track is later used in post-production to improve the atmosphere of a film.

Bit-rate – In digital multimedia, bit-rate refers to the number of bits used per second to represent a continuous medium such as audio or video after source coding (data compression). A larger bit-rate results in better image quality but produces a larger file size.

Cine Log2 gamma – Is a gamma curve with a wide dynamic range designed under the assumption that footage will be recorded under a high-luminance range, such as bright day exterior. It has characteristics that resemble scanned film, producing video files with a cinematic aesthetic

Colour grading – The process of changing the appearance of footage by manipulating the attributes of an image related to colour and exposure. Standard methods include changes in image exposure, contrast, tint, saturation, and levels of black and white points.

Colour subsampling – The practice of encoding images by implementing less resolution for chroma information than for luma (lighting levels) information, taking advantage of the human visual system's lower acuity for colour differences than for luminance.

Colour scopes – During the process of colour grading, scopes help the artist monitor the results by quantifying the results of the image manipulation process.

Dailies – Dailies are the unedited footage for a movie or TV show that is collected at the end of each day for viewing by select above-the-line members of the film crew.

GUI – Graphical user interface is a type of user interface through which users interact with electronic devices via visual indicator representations.

Focus pull – a.k.a. Rack focus is the filmmaking technique of changing the focus of the lens during a continuous shot. When a shot ‘racks’, it moves the focal plane from one object in the frame to another.

Footage conforming – The process of replacing lower-quality media (proxy files) in an edit with higher-quality media, usually original high-quality footage.

Footage bandwidth – Bandwidth of the footage is the maximum data-carrying capacity of a transmission medium and measured in bits per second (bps), Megabits per second (Mbps) or Gigabits per second (Gbps).

Form factor – A camera form factor describes the type of body the camera has as well as the layout of the controls, monitoring surfaces, and mounting points for accessorised equipment.

H264 – H264 or MPEG-4 Part 10 Advanced Video Coding, is a video compression standard based on block-oriented, motion-compensated coding. It is currently one of the most efficient and popular codecs on the Internet.

Histogram – A histogram is a graph that displays the luminance and tones of a video or photography. Most commonly, histograms show the shadows and highlights or levels of red, green, and blue tones to provide more accurate monitoring of recorded footage.

Legacy media – Old media, or legacy media, are the mass media institutions that predominated prior to the Information Age; particularly print media, film studios, music studios, advertising agencies, radio broadcasting, and television.

Lens aperture – Aperture refers to the opening of a lens's diaphragm through which light passes and exposes the camera sensor or film. It is calibrated in f or t stops and is generally written as numbers such as 1.4, 2, 2.8, 4, 5.6, 8, 11, 16, and 22. A smaller number indicates a larger opening.

Lens periscope system – A periscope lens is an arrangement of lenses inside a long tube, with two 45-degree angled lenses at each end. Similarly like in a submarine, the arrangement of a smaller focusing elements along the tube saves space and provides a true optical zoom within a smartphone body, but at the same time, increases the manufacturing cost.

Lighting: high and low key – High-key lighting reduces the lighting ratio in the scene, meaning there's less contrast between the darker tones and the brighter areas. Alternatively, low-key lighting has greater contrast between the dark and light areas of the image with a majority of the scene in shadow.

Motion interpolation – Motion interpolation or motion-compensated frame interpolation is a form of video processing in which intermediate frames are artificially generated by calculating the difference between the existing ones. Motion interpolation is commonly used to compensate for the lack of display resolution, and for artificial slow motion effects.

MP4 – An MP4 file is a multimedia file that stores a movie or video clip and may also contain subtitles, images, and metadata. It is one of the most common file containers used for distributing video online, whether it be streaming or sharing videos.

Product placement – Form of advertising in which branded goods and services are featured in a production that targets a large audience.

ProRes HQ codec – Is a high quality, lossy video compression format developed by Apple Inc.

Video resolution – The number of pixels contained in each frame. The most common resolutions include a standard definition such as 720x576, full high definition 1920x1080, and 4K, 3840 x 2160.

XR Technologies – XR abbreviation stands for extended reality, a term referring to all real-and-virtual combined environments and human-machine interactions generated by computer technology and wearables. That includes augmented reality (AR), virtual reality (VR) and mixed reality (MR) environments and visualizations.

Acknowledgments

I would first like to thank my supervisors, Dr. Claire Henry and Dr. Ian Goodwin, for their support in this research. Developing a one-page proposal into a doctoral thesis requires considerable amount of time, and I wouldn't be able to achieve my goal without their wisdom to keep me on track. I also need to acknowledge many stimulating discussions I held with my wife Sanja while looking to refine research arguments and my young daughter Nastasia's sacrifices for listening to numerous readings over the past four years.

In addition, I would thank my mother and father, who raised me to be thorough and persistent, my sister for her unwavering support in my various endeavours, and our New Zealand whanau, Serge, Christopher, Kylie, Amanda, and Mira for helping us in growing New Zealand roots.

I cannot forget professors from the Faculty of Dramatic Arts in my home country of Serbia, whose knowledge enabled me to express myself as a multimedia content creator on three different continents, as well as the creativity of colleagues in Serbia, the United States, and New Zealand that inspired me to refine my skills. Here I would also add my thanks to artists featured in thesis artefacts; cast and crew of *Loss* (2019), Culture Embassy ensemble, Mr Russel Henderson and Massey COCA team led by James Coyle for giving us access to their flagship Wellington studio for artefacts filming.

Finally, I would like to thank an idle New York City Uber driver, whose name I don't know, but his actions inadvertently inspired this research. While he enjoyed watching 60's classic *Lawrence of Arabia* (1962) on a 4.7-inch smartphone screen parked next to the entrance of the TV station where I worked, he moved me to re-evaluate my judgments on what is considered a professional recording device, appropriate screen orientation, and distribution outlet for work of professionals that eventually grow into this research.

Abstract

Contemporary transformations in the aesthetics associated with online vertical content are a prime example of remediation between mobile video recording technology, diversification of online distribution platforms, and creators' enthusiasm to explore new video forms in exchange for higher audience reach and engagement. By deconstructing texts available on YouTubeVEVO and Netflix through the lens of Applied Media Aesthetics, this research examines the choices that professionals make while they utilise conventions associated with online vertical video and its consumption on a mobile screen. Additionally, through interviews with experienced practitioners, I survey the principal differences in production of widescreen and vertical content, and look at potential benefits and limitations of a smartphone as a primary recording device in the professional production workflow. Lastly, as I plan, film and edit six original artefacts, I reflect on the affordances of a vertical frame through the eyes of a cinematographer on set and editor in post-production. The mixed methodology provides a robust framework for examining the principal benefits and limitations of a vertical frame in a professional production workflow and for answering the key thesis question: how do framing, filming technology, and choice of distribution platforms influence media professionals' aesthetic decisions while creating online vertical content? Throughout the research, I re-examine assumptions about acceptable screen aesthetics commonly held by media professionals and provide insights into the remediation of widescreen content for a vertical frame.

Keywords – Video, vertical, tall frame, smartphone, YouTube, Netflix, music video, social media, media aesthetics, new media, moving images.

Chapter One: Introducing the Thesis

This project was inspired by empirical observations while I worked as a multitasking content producer in the TV and independent film industry. My career in television and film started in the period of large-scale transition from analog to digital workflow, the same year Lev Manovich released his seminal work *The Language of New Media* (2001). From that point, over the past two decades, I have followed the written discourse on the topic of new media and the aesthetics of moving images while witnessing a shift in co-workers' perceptions of vertical content as a 'professionally' acceptable form of expression. For my career, 2016 represented a watershed moment when I realised new distribution channels and portable recording devices would change professional workflows for good. Even traditional TV stations run by the baby boomer generation, such as CUNY TV, where I worked at the time, could not ignore the inevitability of adapting the traditional TV program to various social media channels. I also noted the footage created with specialised audio/video gear, in-studio spaces worth millions, became contested by amateur segments produced on portable devices. For that reason, production teams with decades of industry experience had to accept mobile video as another essential distribution channel due to its immense reach and ties to the smartphone as a ubiquitous recording device.

As someone interested in learning more about new visual conventions, I looked for texts specifically designed for a smaller screen and online distribution. That was my first contact with a vertical format. During the next several years, I experienced the evolution of vertical video from a form of visual communication labelled as 'amateur' into another customary professional practice that profoundly impacted the language of moving images. Rapid developments in video recording technology, particularly the advancement of smartphone cameras and the diversification of online distribution outlets in opposition to legacy media[x], inspired the central thesis question:

How do framing, filming technology, and choice of distribution platforms influence media professionals' aesthetic decisions while creating online vertical content?

Despite the existing body of work on the phenomenon of vertical video and smartphone filmmaking, there was insufficient research in professional practices related to the synthesis of vertical videos and films, nor discussion on the level of smartphone footage integration in traditional post-production workflows. Liv Hausken, in *Thinking Media Aesthetics: Media Studies, Film Studies, and the Arts* (2013), conducted interdisciplinary research aimed at exploring contemporary media practices to make cross-disciplinary theoretical and analytical insights into technical mediation (Hausken, 2013). Others like Miriam Ross and David Neal in 'Vertical film' (2014), 'Vertical 2015' (2015), and 'Vertical Videos from User-Generated Content to Corporate Marketing' (2018), offer a comprehensive overview of the expansion and evolution of the format. Canella in 'Video Goes Vertical' (2017), examined the impact vertical video has had on the production of local TV news, whereas Kathleen Ryan in 'Vertical Video: Rupturing the Aesthetic Paradigm' (2017) discussed persistent popular culture critiques of 'vertical video syndrome'. In early works like David Bolter and Richard Grusin's *Remediation: Understanding New Media*, the authors define remediation 'as the 'formal logic' by which media refashion prior to new media forms' (Bolter and Grusin 1998, p.45).

Although highly relevant for examining the vertical format's evolution, these studies do not discuss how amateurs' adoption of the form influenced professionals' workflow, nor how the implementation of software back-end systems allowing full screen playback or pre-set sequences in non-linear editing programs affected further adoption of tall aspect ratios into the workflow of high-budget projects. These are the pertinent facets for the production and final look of footage that I investigated in this research:

- What are the dominant aesthetic elements of current high budget vertical music videos and trailers?

- What influences the key aesthetic decisions of film and video professionals as they explore the expressive potential of a tall frame?

- What are the benefits or limitations of a smartphone as a primary recording device for the filming of vertical content, in contrast to the use of DSLRs, SLRs or large camera bodies?

- How different is a sequence of moving images created specifically for a smaller portable vertical screen, compared to larger TV and cinema canvases?

- How influential is the smartphone's size in the determination of screen orientation?

- And lastly, to what extent did availability of in-browser software such as Grabyo, and back-end updates of social media platforms (allowing full screen playback), pave the way for broader adoption of vertical format in professional workflows?

The first methodological step was a qualitative content analysis of relevant online texts. Since artists in the music industry are often the first to embrace new technology in the quest for a competitive edge to promote their content (Walzer, 2017, p.6), I looked into the characteristics of the most popular online vertical music videos available on the YouTube VEVO channel. To diversify the sample and include cinematic material, I also looked into the contemporary style of Netflix vertical trailers. Both platforms are focused on online viewers and accommodate a variety of genres. Netflix is available in more than 190 countries, while YouTube has an audience reach of two and a half billion people. A group of dominant visual features is one of the hallmarks of every established genre in the history of moving images, and my preliminary survey found that material on these channels developed a unique set of visual characteristics, making them intriguing for further research. The textual analysis focused on presiding two and three-dimensional components in order to search for dominant aesthetic patterns. Furthermore, while analysing Netflix vertical previews, I also got a data set necessary to understand the creative decisions of an editor tasked with re-purposing widescreen footage into a tall aspect ratio.

The subsequent step in the research involved a set of semi-structured interviews with a sample of TV/film producers, directors, cinematographers, and editors. The conversations provided details into principal differences between horizontal and vertical content production in a professional environment, as well as insights into potential limitations or benefits related to the choice of the distribution platform. The interviews were also a follow up to Gino Canella's investigation conducted in Video Goes Vertical (2017), as I probed the opinion of professionals towards vertical content as an acceptable and relevant form of expression five years later. Finally, being a professional myself, I put the research conclusions to the test by producing vertical content for both online platforms used in the text analysis. The applied component consisted of a production of six artefacts; two vertical (9:16 aspect ratio) and two widescreen, (16:9 aspect ratio) music videos, followed by the editing of a vertical film and associated tall trailer in slimmer 1:2 aspect ratio. The findings from the preceding methodological steps informed my key aesthetic decisions during the production of thesis artefacts, especially the filming style of the videos.

The sentiments of spontaneity and authenticity are often associated with user-generated vertical content, and professional creators wish to carry that feeling by simulating the amateur filming style in scripted and directed content. For that reason, I opted to film the first four items as single-take music videos in a style similar to the key texts I analysed on the YouTube VEVO channel, where the focus of audience attention wouldn’t be guided by editing but rather with camera movement and image aspect ratio. Additionally, while filming vertical and widescreen footage of identical performance interchangeably, I could document benefits and limitations between the two configurations with great accuracy. Correspondingly, while looking to create original items matching the style of Netflix vertical previews, I edited the last two artefacts from original 16:9 widescreen footage into 1:2 frame, documenting and

reflecting on my creative decisions along the way. Equally important, through the entire process of artefact creation, I was able to employ my extensive experience in video production, cinematography, and editing, while reflecting on principal differences, benefits, and limitations of vertical video making. After the literature review in chapter two and chapter three detailing methodology, I analyse dominant 2D and 3D characteristics of relevant texts in chapter four, followed by interviews with media professionals in chapter five. Lastly, through the making of artefacts in chapters six and seven, I reflect on the conclusions from all previous methodological steps as I scrutinise footage re-purposing workflow during the editing of short fiction vertical film and accompanying vertical trailer cut from original widescreen material.

Chapter Two: Literature Review and Background

In order to answer the principal thesis question, I needed to consider where vertical video stands in the contemporary media world; but firstly, what kind of video is defined as a vertical video and what is an appropriate method to analyse it? According to Anne Friedberg in her book *The Virtual Window: From Alberti to Microsoft* (2009):

> A vertical video is a video created either by a camera or computer that is intended for viewing in portrait mode, producing an image that is taller than it is wide. Thus, it sits in opposition to the multiple horizontal formats matching the plane of eyesight and normalised by cinema and television, which trace their lineage from the proscenium theatre and Western easel painting traditions (Friedberg, 2009, p.131).

Media aesthetics is a multidisciplinary and heterogeneous field for research on media technologies, aesthetics, and mediation. Hausken summarises its various influences in the introduction of her book *Thinking Media Aesthetics: Media Studies, Film Studies, and the Arts* (2013):

> One could say that media aesthetics emerged from earlier attempts to describe relations between technology and media, such as media philosophy (Friedrich Kittler), media ecology (Marshall McLuhan), medium theory (Joshua Meyrowitz), mediology (Regis Debray), and critical theory (Walter Benjamin). Today, it is mainly influenced by aesthetic ideas of what has been termed new media (Lev Manovich) and visual culture (W.J.T. Mitchell) (Hausken, 2013, p.29).

The more practical approach to this topic can be found in Herbert Zettl's *Theory of Applied Media Aesthetics*. Since vertical video presents a relatively new phenomenon, alongside academic sources I also cite studies conducted by reputable organisations such as Oxford Reuters Institute, Pricewaterhouse Coopers, Deloitte, Bond Capital, Scientia Mobile and articles from magazines like *Wired, Forbes, Wall Street Journal, The Guardian* and relevant blogs.

Zettl developed 'applied media aesthetics' (hereafter AMA) as a collection of methods and conventions that makes artistic decisions in video and film production less arbitrary, while simultaneously facilitating the analyses of a variety of aesthetic variables. For these reasons, AMA offered a robust theoretical framework suitable for the deconstruction of both the new and old moving images conventions I recognised in vertical videos. The last, eighth edition of Zettl's *Sight Sound Motion* discusses the rapid advancement of digital technologies after the 2000s, and the proliferation of smartphones as portable recording and playback devices, yet, a section on vertical filming considers only the possibility of rotating the device and changing the footage aspect ratio in editing software. That is understandable since the book was published in 2016, when vertical video was only just getting traction as a global phenomenon. However, Zettl's existing in-depth analysis of various traditional aspect ratios, (such as 4:3, 16:9, and 2.35:1) as well as the comprehensive breakdown of two and three-dimensional image parameters available in AMA, offered a strong foundation for my research of vertical framing.

Before various online distribution channels developed back-end support allowing full-screen playback, a vertical video wasn't played in the same aspect ratio as it was recorded, and it was commonly associated with user-generated content. If housed in a widescreen-orientated display, the disparity between recording and playback screens created aesthetically unpleasing pillar boxing of the image. When using this method, vertical video is simply inserted into the centre of the broader screen, so that the top and bottom of its frame coincide with the edge of 16:9 picture space. The leftover screen sides are filled with black bars, and being vertically orientated, these sidebars look like pillars hence the term pillar boxing (Zettl, 2016, p.94). In the same way, if any widescreen video is inserted into the 9:16 aspect ratio, the footage would exhibit thick top and black bottom bars called letterboxing as you can see in figure 1.

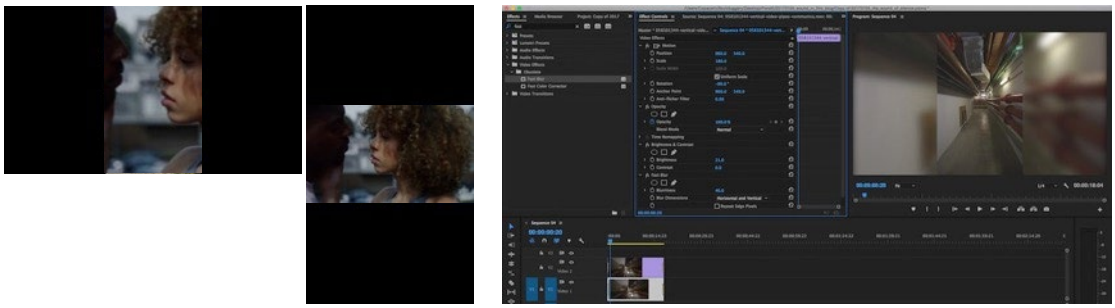

Figure 1: An example of pillar boxing, letter boxing, and image blow-up in Adobe Premiere Pro.

Looking for an alternative to black bars, a third post-production method to fill the screen emerged over the years in the form of footage 'blow-up'. By duplicating and magnifying the same file in non-linear editing software, the editor could substitute black bars with footage that somewhat aesthetically resembled the style of the central frame. The blow-up solution was far from ideal, as the editor essentially created a set of blurred bars, but matching colour, recognisable mise-en-scene elements, and similar direction of camera movements within the composition, formed a better connection to principal footage, in any case, more relevant than large black volumes. All of the listed solutions involved aesthetic compromises that left large parts of the mobile screen occupied with non-relevant material, consequently upsetting audience perceptions of the principal footage. As Ryan remarked, both landscape and portrait videos appeared equally distorted, and equally wrong, when viewed in their non-native orientation (Ryan, 2017, p.11).

One of the best-known early objections to vertical video, *Vertical Video Syndrome A PSA* (2012) created by Glove and Boots, reflected the contemporary mainstream perception of vertical format. This piece is analysed by Ryan and Ross, and often quoted by other scholars interested in the remediation of vertical video. Originally posted to YouTube in June 2012, it garnered more than eight million views, got 130,000 thumbs-up votes, and had more than 5,000 comments before posting was disabled. As Ryan noticed, 'the faux public service announcement hit a nerve' (Ryan, 2017, p.2). Despite its popularity, the original link was taken down in 2017, and today only copies with significantly lower view count can be found (Glove and Boots, 2012). So, why did the authors take down their most popular YouTube segment? The instigators Damien Eckhardt-Jacobi and Vincent 'Vinny' Bova, explain the reasons in their Facebook post *'*A message from the Glove and Boots Creators' (March 2019) by critically reflecting on an eclectic mix of videos available on their channel.

> Our content has always been scattered. We sang children's songs, then shot a zombie in the head, then made a top ten list – attempting to ride whatever Internet wave we could. The YouTube landscape has been chaotic – particularly for channels like ours that sit somewhere between children's entertainment and adult content. As YouTube attempts to limit children's exposure to inappropriate content, channels like ours fall into a weird vacuum (Glove and Boots FB, 2019).

This post offered a remarkable insight into the behind-the-scenes circumstances surrounding online video bloggers in the Internet age. In 2012, the vertical video was an emerging format. For that reason, Glove and Boots used the novelty to draw attention to their channel by reinforcing common stereotypes associated with 'proper filming orientation.' While creating a satirical public service announcement Vinnie and Damien never planned to create the manifesto against vertical video. Their post went viral by chance, and the reasons for that are not part of this research, but the additional exposure undoubtedly drew attention to the developing phenomenon and fuelled already heated debate around it.

The rise in popularity of vertical video coincided with the global expansion of the Internet and a steady decline of TV viewing as the dominant medium for the consumption of moving images. Traditional broadcasters have struggled to attract new users for the past decade, especially among young people, some of whom do not even own TV sets (NY Times, 2018). According to an annual survey by

Leichtman Research Group, in fall 2014, 84 percent of TV households in the U.S. subscribed to some pay-TV service – either cable, satellite or telco (Leichtman Research, 2019). At the same time, fewer than half of TV households had a streaming subscription to either Netflix, Hulu, or Amazon's Prime Video, the three most popular over the top (OTT) platforms in the US in 2014 (Ooyala, 2015). By fall 2018, the percentage of TV households with cable, satellite, or telco service had fallen to 78 percent, while streaming services subscription grew to cover almost three-quarters of all homes in the U.S. (CBS, 2019). As Anna, the youngest TV professional interviewed in my research summarised: 'I worked every day to make TV, but then I didn't watch TV, my friends didn't watch TV, no one I knew really watched TV. So I said to my team, we need to make something on the Internet' (Anna). I will unpack the discussion on aesthetical and technical decisions working media professionals make as they produce vertical content in chapter five, as we talk about the level of its integration into their regular workflows.

In contrast to the declining numbers of legacy media users, the amount of people accessing the Internet via smartphone grew sharply. By 2019, 3.8 billion people had daily access to the Internet (Meeker, 2019, p.4), while 3.5 billion owned a smartphone (MOVR, 2019, p.3). A Deloitte digital report from 2017 found that smartphones were near-ubiquitous among millennials, with ownership in the US, Canada, and Great Britain reaching 94 percent in 2015 (Deloitte, 2017, p.5). American business news channel CNBC estimates that this growth will continue as nearly three-quarters of the world will use just their smartphones to access the Internet by 2025 (Handley, 2019). Statista's study published in February 2022 demonstrates the expansion of the global Internet reach continues to be driven by mobile users. According to this study, in 2021, the number of unique Internet users stood at 4.32 billion, with 90 percent of the population using a mobile device to go online. Mobile ownership and Internet usage are forecast to keep growing in the future, as mobile technologies are becoming more affordable and available than ever (Ceci, 2022). Even Marshall McLuhan could not predict such a dominance of a single medium when he proclaimed half a century ago that 'the medium is the message' (McLuhan, 1964). Faced with the decreasing numbers of audiences using traditional channels for program distribution, television stations, as well as commercial brands, have continued to look for new ways to engage customers through social media, and tried to accommodate their content to online mobile platforms. Being a unique filming style, primarily associated with consumption on a smartphone, the popularity of vertical video within mobile audiences did not pass unnoticed.

The highly contested social media market, television industry, and commercial brands were not the only ones noticing the online popularity of the new format. The equally saturated purview of short films also benefited from a new kind of storytelling. Established in 2014, The Vertical Film Festival (VFF) in Katoomba, New South Wales/Australia, was conceived to encourage the exploration of vertical film and video. The VFF was the first worldwide competition focused only on the vertical format and has since become a biennial event with public screenings in suitably vertical venues (VFF). At the time of this writing, there were no movie theatres explicitly designed to screen vertical format yet, so the architecture of appropriate, makeshift venues had to correspond with the shape of a tall frame. St Hilda's Church in Katoomba, with a nave significantly taller than wide, became one of those convenient locations used for screening of vertical films (figure 2). Another creative collective, the Mobile Innovation Network and Association (MINA), now in its 11th year, hosts the longest-running international film festival dedicated to mobile & smartphone filmmaking (MINA). An overview of VFF and MINA archives from 2014 onward displays the steady progress in the technical complexity and aesthetical refinement of submitted vertical films. But what about vertical content before it became popular enough to get enough attention from early adopters with professional skills willing to explore the new format?

Figure 2: St Hilda's Church in Katoomba, Australia.

Tracing back the history of vertical content, the application that primarily influenced the early rise in vertical video popularity is Snapchat, launched in 2011 (Neal & Ross, 2018, p.153). The company thoroughly promoted the idea of improved user-experience through vertical framing by developing a

ground-up application that favoured the format (Snap INC). From its inception, Snapchat advertised the vertical format as a new mode of communication targeting mobile users, and it quickly drew the attention of troubled broadcast outlets searching for new ways to engage the audience. Steven Braband, director of ESPN digital video operations, describes this remediation between television and mobile video conventions in a Forbes interview from 2018:

> Vertical video has changed our editing process. Before launching SportsCenter on Snapchat, virtually every highlight we created would first use the traditional 'game angle' and then replay angles on some of the biggest plays. Now, if you look at our SportsCenter Snapchat show, we're often going straight to the replay angle of a big LeBron dunk, and that's all we need, (Forbes, 2018).

Even though the platform was initially built to tell short stories, Snapchat quickly understood the future potential, and just a year later introduced a longer format named Snapchat Stories. In contrast, YouTube was substantively more cautious. After an announcement that it would improve support for the vertical format in 2013, it took YouTube two years to update the application code. Soon after updates were finished, the company introduced curated lists of the most popular vertical music videos, and other catalogues of vertical content sorted by genre. Throughout 2015, other major social media platforms such as Facebook, Twitter, and Periscope (acquired by Twitter in 2015 before public launch) also embraced the format. The amendments in the application code allowed full screen playback of vertical content on their platforms, without the need for pillar boxing or footage blow-up. As more digital natives began using their smartphones as primary device for Internet access and entertainment, the popularity of vertical content continued to rise. In the 2015 annual Internet Trends Report, Mary Meeker noticed that the percentage of vertical videos people watched online jumped from five to 29 percent between 2010 and 2015, matching the trend with the growth of Snapchat, and to lesser extent, other social media applications such as Instagram, Vine, and Musically (Meeker, 2015, p.68). In 2018, Netflix joined the group of late adopters, and in the years after, it introduced a new way to promote vertical content by redesigning its mobile and desktop application to prominently feature 'vertical previews' (Netflix, 2018).

Instagram took a different integration approach by teaming up with brands already popular on the platform. National Geographic recognised new expressive possibilities of a tall frame combined with the audience engagement potential of mobile video. The company announced it would air the final episode of *One Strange Rock* on Instagram, as soon as the platform's new long-form video hub Instagram Television (IGTV) went live (Instagram Press, 2018). In June 2018, users could watch the full episode of the critically acclaimed documentary series hosted by Will Smith and executive produced by Darren Aronofsky from National Geographic's Instagram account for free. With more than 100 million followers, National Geographic was one of the first major media companies to take a made-for-TV video and adapt it for online vertical distribution (Nat Geo Press, 2018).

The next step in the integration of vertical content into the broader media landscape came with the launch of Grabyo. Being a cloud-based platform, Grabyo removed the traditional complexities of professional video production and distribution processes. While accessing the suite from a browser, producers can remotely create digital and social video from anywhere in the world, using only a laptop and Internet connection (Grabyo, 2020). In 2021, this system was embedded into popular social media applications such as Facebook, YouTube, Snapchat, and Twitter. Grabyo is also regularly used by news agencies like Thomson Reuters and some of the world's largest broadcasters like Sky Sports, ITV and BT Sport (Grabyo, 2021).

Despite expanded integration into online platforms and increased versatility of non-linear editing software, the adaptation of video or cinematic sequence into new aspect ratio is not a simple process, as even brief analysis reveals apparent differences in the construction of two- and three-dimensional space available in wide and tall frames. For example, the area necessary for composition of over-the-shoulder, or two-shot as a means for audience spatial orientation is considerably reduced in vertical composition. To frame the figures in the shot together, both subjects have to stay closer than they usually would in any widescreen composition. Subsequently, the forced proximity of the bodies will

affect the psychological closure of the composition, and influence how the audience perceives the relationship between them. In addition, aesthetic characteristics of the image, such as perspective distortion and lens perspective compression, produced by various wide, normal, and telephoto lenses will render different results if footage is housed in a narrow frame. However, since the upright human body presents a shape significantly taller than wide, the vertical frame offers new possibilities for constructing exceptionally balanced compositions, particularly if a story is built around a single protagonist. Additionally, an image frame with a longer vertical edge offers new possibilities for Z-axis[x] articulation and exploration of depth within the shot. Then, how different are blocking, cinematography, and editing techniques when operating in a vertical frame compared to a widescreen frame?

Through his analysis of how new media relies on older forms and how it breaks with them, Manovich in *The Language of New Media* (2001) looks into new conventions by placing them within the history of modern visual and media cultures. He also poses another question highly relevant to this research: 'How do old media conventions and techniques such as rectangular frames, mobile viewpoints, and montages operate in new media?' (Manovich, 2001, p.34). Following a similar logic, this research found some long-established professional practices prove to be a good fit for vertical video, while others all but disappear due to the nature of a slimmer frame and editing convention associated with mobile media. For example, traditional methods of continuity editing commonly known as 'classical Hollywood editing,' were used less in the online videos because the editor often lacks multiple camera angles necessary to build 'an invisible cut' between the shots in the sequence (Harper, 2011, p.50). A few other techniques such as deductive editing, slow-paced camera dollies and very wide establishing shots are rarely used in online vertical videos, as they do not fit well in short running time nor small screen. In contrast, the editing techniques such as montage developed to condense diegetic time, as well as the direct address method used to encourage engagement, prove to be an excellent fit for the audience's short attention span and the desire for interaction commonly associated with online videos. Nathan Peterson from Forbes Communication Council sums up this evolution of editing in his comments on sports highlight videos.

> The vertical format inherently requires a closer, cropped, more refined take. With society used to the traditional widescreen format, it can be challenging to share the same depth in terms of the context of a short story. However, the vertical orientation offers immediacy, an intimate view and interactivity as the consumer hold their touchscreen smartphone in the upright position (Forbes, 2018).

'Refined take', with less ambiguity and simplified mise-en-scene elements, summarises the principal difference between professional vertical music videos on YouTube VEVO and their horizontal counterparts, which I analyse in more depth in chapter four.

Even though vertical video offered a new way to engage online audiences and immerse the spectator in the story, its integration into the professional workflow of legacy media was far from seamless. In 'Video Goes Vertical: Local News Videographers Discuss the Problems and Potential of Vertical Video' (2017), Canella examines what impact vertical video has on the production of local TV news through interviews with 15 media professionals, current and former local television news videographers and editors. Canella underlines that study participants predominantly defend a widescreen interpretation of the event as 'professional' for most of the time and label vertical as an 'authentic' user account, a sentiment that according to interviews I conducted in chapter five is still present in 2022. 'To distinguish their work as a professional and brand vertical video as 'amateur', interviewees relied upon their learned knowledge of aesthetic principles, workplace experience, and training' (Canella, 2017, p.8). Canella also points out: 'a professional habitus is naturalised and maintained in the workplace but, interestingly, may first be established in journalism schools and colleges'(Canella, 2017, p.9). As chapter five of this research shows, the antagonism of working media professionals towards vertical content softened since 2017, but it is still recognisable in a sub context of a strong personal preference towards traditional widescreen aesthetics. One statement resonates particularly strongly with the overarching feeling associated with vertical video at the time of Canella's investigation in 2017. A younger media professional stated that her:

> lack of technical training could, in fact, be an advantage. She shoots the video wide-screen if she wants to get more of the scene. But, because she does not feel the need to adhere to professional 'guidelines', she is able to find creative applications for vertical video in her work (Canella, 2017, p.11).

Canella further stresses that new technologies like smartphones and mobile applications appear to replicate the distinctions journalists draw between their work and the work of an untrained public. Unlike the concept of preferred framing suggested by online distribution platforms in pursuit of a better user experience, I would claim this argument is mainly obsolete in 2022. The properties of video footage that contemporary smartphones can record is now at a level where even professionals need to consider the benefits of a smaller recording device, against the price and limited mobility of broadcast-grade equipment. As Canella concluded:

> some video professionals working in local TV news will likely find creative uses for vertical video on their station websites and mobile apps, but regardless of how professionals feel about it, consumers will continue shooting and sharing vertical video via mobile apps and devices designed to support the format (Canella, 2017, p.16).

In the years that followed, it was clear that millennials' viewing preferences quickly overpowered the initial backlash from professionals. Mobile viewership significantly outperforms TV, laptop, and desktop viewing when it comes to alternative ways to consume video for this age group. The only exception to this trend is longer drama forms and reality TV shows (Verizon, 2018). The existence of vertical TV sets like Samsung Sero, demonstrates major brands' desires to appeal to Gen Z and Millennials as primary aggregators of vertical content. According to Observer, those two age groups create 43 percent of all videos on Instagram, Snapchat, and TikTok in portrait mode (Harmon, 2020). By offering a 43-inch TV set that automatically syncs to a smartphone and can switch orientations easily (figure 3), Samsung wishes to engage users' preference for vertical format in recording and playback time (Samsung, 2019).

Figure 3: Samsung Sero (2019).

Despite the existence of Samsung Sero, such specialised TV sets still occupy an insignificant portion of the global market. The reason for the long-term dominance of a rectangle might be the fact the majority of already produced video, film, and computer screens are horizontally oriented, and knowing that the following question comes to mind; is there a direct relationship between traditional screen orientation and the horizontal position of our eyes? James J. Gibson explains this correlation in *The Ecological Approach to Visual Perception* (1979). 'When standing still, normal peripheral vision is broader on a horizontal plane (about 180 degrees) than on a vertical one (about 140 degrees), and expressed in a rectangular aspect ratio, this relationship between X and Y edges comes to approximately 4:3' (Gibson, 1979, p.206). However, while moving, this broad field narrows down significantly as our peripheral vision diminishes rapidly.

At this point, it is important to stress, the shorter horizontal axis of vertical image frame or smaller mobile screen does not immediately mandate the necessity for a stationary camera, and Gibson also explains why.

> Even if a person is standing still and the screen is small enough to be reflected fully at our retina, we cannot see all of its content by simply staring at it. Eyes must constantly cut from one spot to another in a picture before making sense of what we see. Such eyeball jumps are called a saccadic movement (Gibson, 1979, p.209).

The author explains that the combination of saccadic movements and temporary fixation constitutes the act of scanning and subsequently forms an image in a person's brain. Even before being able to accurately measure the saccadic movements with the help of sophisticated technology, 1970's film director Edward Dmytryk, explored how eye movements can inform the editing rhythm and improve

the sense of visual continuity in theatrical projections. Dmytryk described an appropriate moment to cut by estimating the time an average user needs to shift their eyes from one point to another.

> To make the cut, we fix the frame in which the actor's eyes have 'frozen', add three or four frames more to give the viewer time to react and move his eyes as he follows the actor's look, at which point the cut is made. (Dmytryk, 1986, p.444).

The time space of three to four frames equals 125-167ms at 24 frames per second, and it is similar to the minimum time taken to perform a saccadic eye movement found by authors like Smith using sophisticated eye-tracking computer systems decades later (Smith, 2013, p.9). Going deeper into the characteristics of human perceptions exhibited while watching moving images, Smith and researchers like B. Fischer & E. Ramsperger, he quoted in his study, found an average audience member can scan 2-5 locations within a screening canvas during one second depending on the image luminance levels, the dominance of movement in mise-en-scene and objects saturation levels (Smith, 2013, p.10). Equally important, virtually limitless combinations of three aforementioned elements dominate human perception significantly more than the shape or size of the screen. To summarise, a large viewing canvas or widescreen will not eliminate the scanning process, but as Zettl observed, 'it will show enough of a scene in peripheral vision that the spectator becomes aware of what is going on at the sides of the screen before deciding to look at it' (Zettl, 2016, p.87). In other words, the energy of a large screen can improve the viewing experience, but the expressive potential of theatrical against the mobile screen or wide versus vertical aspect ratio are the same from the point of human perception.

Looking back at the history of moving images, it is understandable that the industry and the audience are more comfortable with the familiar aspect ratio of widescreen and the language of moving images commonly associated with it. After all, the cinematic conventions such as the rule of thirds, the embedded meaning of camera angle and height, as well as the editing principles for the control of diegetic time and space were refined by Edwin S. Porter, Georges Méliès, David W. Griffith, Sergei Eisenstein and other pioneers of a film more than a hundred years ago. In many ways, early travelling filmmakers were equivalent to today's online bloggers and influencers, since they used the best technology of the era in conjunction with any available viewing platforms, such as travelling and makeshift theatres. One of the devices enabling such versatility at the pioneering age of cinema before the twentieth century was the Cinématographe (figure 4). Made by the Lumiere brothers in the 1895, the Cinématographe was a hand-cranked small camera, nothing like a cumbersome DC powered Edison's Kinetoscope. For the creative people of the era, that meant the camera could be easily transported and operated by a single person in the field, and equally important, by flipping the lens, Cinématographe could project the developed film on any clear surface (Britannica, 2019). Compared to today's advanced smartphones, such features transformed Cinématographe into a portable, multipurpose recording and playback device utilizing the newest 35mm film technology of the era.

Figure 4: Lumiere Cinematograph.

The capabilities of a Cinématographe made moving images very accessible for a broad audience, but before the expansion of theatres, the aspect ratio of the projection depended solely on the creators' preference and conditions at the screening location. Due to the absence of technical specifications for the projection and venues, a travelling filmmaker could display their work on any clear surface, making the standardisation of aspect ratio almost impossible in the early decades. Some of those similar traits we can attribute to smartphones and online vertical video today, as new aspect ratios presented both opportunity and challenge for a mobile screen's storytelling aptitudes. The technical improvements, in conjunction with the simplicity of smartphone use as a recording and playback device, paved the way for a proliferation of user-generated content and vertical video despite the fact there is no global standard for smartphone screens. At the same time, the possibility of accessing – and misunderstanding – online videos produced in a foreign language became even more apparent as the Internet blurred countries' borders. Subsequently, if content creators want to make a video understandable to a global

audience without captions and additional descriptions; they must abide by well-established traditions of the language of moving images, as those conventions already bridge the language barrier. Nevertheless, no matter how much technology is refined, the fundamental principles of composition and key editing techniques for building diegetic space and time control are similar to the early days of cinema and recognisable in vertical video today.

The evolution of aspect ratio did not stay limited to a widescreen and vertical rectangle. By 2019, 9:16 (YouTube videos) and 1:2 (Netflix trailers) aspect ratios became a common site on the Internet, together with others like 1:1 square (Instagram) and circular frame (Snapchat Spectacles). Despite being deeply knotted in remediation with older widescreen formats, most of the new aspect ratios stayed platform-centric and did not become a new dominant standard for moving images. So if there is no particular shape preference for the viewing screen from the point of human perception, what other factors made widescreen such a dominant format? Ted Hovet in *Persistence of a Rectangle* (2017) reviews the challenges authors, producers, and distributors encountered during cinema's shift from more square 4:3 to widescreen aspect ratio. He notes it took the cinema industry two and a half decades to fully adopt widescreen, while filmmakers perfected the new ways of cinematic storytelling offered by a larger canvas. Hovet also quotes Carl Lewis Gregory, who argues 'that behind the teleological history and advocacy of standardisation's technical and economic advantages lies a powerful argument for a norm that would admit deviations only under extreme pressure' (Hovet, 2017, p.140). In the case of online vertical video, that pressure came during 2017 with a proliferation of user-generated content and a push for broader adoption of a tall frame from well-funded platforms Such as Facebook, YouTube and Snapchat, aware of its engagement potential on social media (Business Insider, 2019).

The issues of unstandardised framing became less visible today since most web pages and electronic displays can screen any footage in the original aspect ratio without cropping. On the other hand, a creator's challenge to design a composition able to convey the same message without ambiguity in two or three alternative aspect ratios became increasingly relevant again with a profusion of social media channels favouring particular footage shapes. Being of substantial influence on the planning of mise-en-scene, post-production workflow, and distribution possibilities, the choice of aspect ratio is one of the key image parameters. The decision regarding the final aspect ratio of the image is equally vital in multimillion blockbuster projects and music videos targeting the mobile audience, as you can see from figures 5 and 6.

Figure 5: A selection of crop masks for various aspect ratios of theatrical projections, superimposed on a screenshot from Dunkirk (2017).

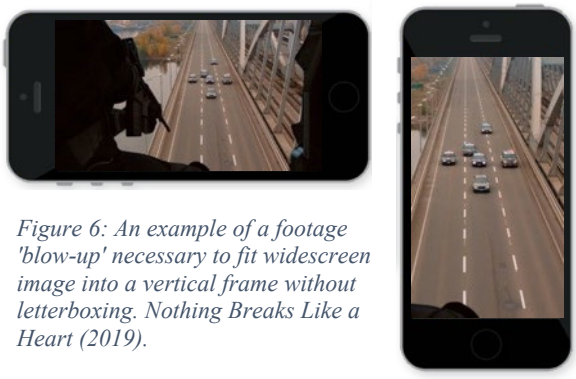

Figure 6: An example of a footage 'blow-up' necessary to fit widescreen image into a vertical frame without letterboxing. Nothing Breaks Like a Heart (2019).

A director commonly had to include additional mise-en-scene elements on top and bottom of the original composition to create 'screen safe' areas, knowing they may or may not be included in the projection. For that reason, the director needs to ensure elements close to the composition edge are not vital for understanding a plot but are still pertinent to the action if they appear on the screen. From figure 6, you can also see how cropping of widescreen footage, which is necessary to project the image without letterboxing on a vertical screen, produces a significantly different composition than the original. I will further elaborate on these creative compromises in chapters four and seven, but if I set aside the issue of unstandardised projection and cropping for now,

what about personal preference for vertical framing? Is it fair to say that shorter forms such as private messages and random clips from everyday life, breaking news, trailers, and music videos are acceptable in any orientation, and longer fiction or documentary formats are not?

Ryan in 'Vertical Video: Rupturing the Aesthetic Paradigm' (2017) examines in depth the personal aesthetic desirability of vertical videos through Applied Media Aesthetics. Because it utilises the same theoretical framework, Ryan's article is particularly relevant to this research with a one key distinction: one focuses on user-generated and another on professional content. Rupturing the Aesthetic Paradigm' explores the feasibility of a new vertical video aesthetic, the notion of 'proper' framing in a horizontal plane, and the reasons that a vertical video aesthetic may be desirable for smartphone screens. In her conclusion, Ryan summarises:

> Users don't make the vertical video because they are somehow ignorant of traditional video orientation. They don't make a vertical video because they are too lazy to rotate the phone 90 degrees into a 'proper' video orientation. Users make the vertical video because it makes sense within the technology. In doing so, they are transforming the aesthetic of the moving image (Ryan, 2017, p.17).

I agree with Ryan, but I would add that at the time of her writing in 2017, vertical video was still in the process of rupturing the aesthetic paradigm, while in 2022 it occupies a more mainstream position due to application support developed in the meantime, a significantly larger volume of texts available online, and as this research shows, new aesthetic conventions developed in vertical format through professionals' increasing use of it. Furthermore, the perpetual evolution of technology, introduction of new distribution channels, the proliferation of mobile video and the development of XR[x] technologies created the impulse for more practice-led research in academia, that in return, gave more insights into conventions of mobile video recording and vertical content within. Relevant examples in the region can be found in the Australian Screen Production Education & Research Association (ASPERA) works. The research films, like *Remembering Hiroshima* (2017) and *How Many Ways to Say You?* (2017), explore how smartphones may be used for auto-ethnographic research while providing artists/filmmakers with the means for spontaneous and simple visual data collection (ASPERA, 2017).

Another pair of researchers, Ross and Neal, who followed the rise of vertical video, produced a survey of the state of vertically framed moving imagery' in two reports: 'Vertical film' (2014) and 'Vertical 2015' (2015). Both documents offer a review into an intriguing variety of genres, styles, support applications, and a growing number of endorsements in addition to online distribution outlets able to support vertical content. However, as in the case of Ryan's essay, these studies cover formative years of vertical content, and due to a rapid expansion of online video on the Internet, present-day lists had to be curated with a narrower set of parameters. Nevertheless, these older catalogues still offer a valuable overview of the evolution and chronology of vertical format adoption. In 2014, Ross produced another article, 'Vertical Framing: Authenticity and New Aesthetic Practice in Online Videos', that looked deeper into the trend of shooting in a vertical mode, and the material markers of 'authenticity' this mode appears to lend to its audio-visual content (Ross, 2014). Besides analysing contemporary trendy user-generated videos, this paper also looks back at older examples of professionally created vertical content like Paolo Gioli's vertical film*s* *Stenopeico* (1973/81/89), *Commutazione Con Mutazione* (1969), and *L'operatore Perforato* (1979). The article also covers the conceptual works of Bill Viola's *The Messenger* (1996) and *The Crossing* (1996), vital for the exploration of vertical aspect ratio through critically acclaimed artwork (Viola, 2020). In combination with the creation of unique texts, the author explores the challenges these experimental films pose to understanding traditional compositional strategies, the choice of topic, and the delivery format. Since there is a significant difference between the energy of a smartphone screen and a gallery wall as a place for artwork consumption, I concur with Ross that the choice of subject, final delivery format, and viewing platform has a large role in determining production parameters, but I would suggest another attribute that influences the smartphone's filming and playback preference even more strongly, that is a device's form factor[x].

Lastly, while refining the principal thesis question and developing an appropriate methodology to answer it, I had to acknowledge the fast-changing dynamics of the field I was researching. For that reason, I situated this study in formal aesthetic tradition to keep a stronger focus on the analysis of key texts produced between mid-2017 and the end of 2019, as well as their hallmark aesthetic characteristics related to the evolution of vertical video at that point in time. Following Richard Dyer's logic summarised through the closed-system, open mind approach, I had to accept 'it is never possible to do everything. Most of the time one has to put on hold crucial aspects of a phenomenon that one has not time (or perhaps inclination) to address' (Dyer, 1988, p.11). A narrower focus also improved my ability to get a deeper understanding of aesthetic and technical decisions media professionals make while producing vertical content. At the same time, I acknowledge technical and aesthetic decisions can be influenced by a person's private and work environment, as well as broader cultural settings as John Thornton Caldwell discusses in *Production Culture: Industrial Reflexivity and Critical Practice in Film and Television* (2008). For example, the encouragement of team or solo productions in a work environment affects the range of topics a content creator chooses, while access or lack of it to the wide range of portable audio-video gear directly relates to an employee's field production versatility. However, despite acknowledging its relevance, I knew from personal experience that looking into the affordances of production cultures would open Pandora's box of questions that would take me outside the scope and focus of this research.

Following the same rationale, I did not engage in the equally vital discussion on personal representation associated with vertical framing in the online environment, despite the fact the analysis of vertical video texts provided some valuable insights on that topic. My research shows that the vertical frame works aesthetically better combined with a simplified mise-en-scène and less complex compositions. In a broader sense, such filming style invites a sense of individuality, encouraging the neoliberal sentiment that one has great control over personal representation in an online space. Furthermore, the way the scene is framed encourages the audience to read the scene in ways that have cultural significance, for example, in the representation of body image.

The analysis of texts such as the music videos discussed in Chapter Four could provide more insight into tall frame implications on representation, but once again, following Dyer's rationale, I had to acknowledge the limitations coming from the choice of the objectivist research approach. The affordances of vertical framing and deconstruction of dominant 2D and 3D image parameters fit well into the objectivist framework, since ‘The objectivist see the truths of science as facts discovered in the natural world and technical practices as things imposed upon practitioners by apparatuses, while historicist research approach depends significantly more on cultural paradigms’ (Dyer, 1988, p.11). Nevertheless, while acknowledging the proliferation of vertical video has a profound impact on representation, as it encourages individuality and influencer/celebrity culture, to debate those issues, I would have to design a different methodology, one that would compromise my ability to answer the principal thesis question.

Chapter Two: Smartphone Size and Comfort of a Single Hand Grip

This brief overview of cell phone to smartphone evolution illustrates the progress of mobile technology, with a particular focus on device dimensions that affect a user's preference for single or two-hand operation and the quality of camera. In searching for the dominant trend of mobile and smartphone size, I compared more than 200 flagship and most popular models produced by major global manufacturers from 1983-2021. The analysis revealed after a period in the late 1990s when small size and weight were fashionable; mobile phones started growing back in size over the past fifteen years, settling on 6.7 inches (169mm) screen diagonal for the largest contemporary flagship models such as iPhone 13 Pro Max and Samsung Galaxy S21 Ultra. Once batteries became powerful enough to offer longer operating times, and other components shrunk to the point where weight stopped being an issue, manufacturers started competing by offering sophisticated video recording capabilities and a larger storage space. That evolution is apparent in specification comparison between Motorola DynaTAC (1983) and Sharp J-SH04. Being the first commercially released cell phone, Motorola DynaTAC (1983), offered 35 minutes

of talk time and eight hours of standby in a device 25cm high, weighing 1130g. Seventeen years later, the first commercially available mobile phone with an integrated camera, the Sharp J-SH04, was launched in November 2000 in Japan. It weighed 74g and had a single rear-facing camera capable of delivering 0.11 megapixels photographs. After another five years of technology development, Nokia N90 had a twenty times larger pixel count on its sensor, Carl Zeiss optics, autofocus system, standard definition[x] video recording limit of two hours, and an LED flash in the device weighing 173 grams (Hill, 2013). Here is a summary of a trend I observed in the evolution of the mobile phone weight and size; a full list is available in appendix #1. Dimensions in figure 7 are listed in millimetres.

Name	Released	Width	Length	Thickness	Weight/g
Motorola DynaTAC	1983	42	250	75	1130
Motorola StarTAC	1996	55	94	19	95
Ericsson T28	1999	57	97	15	83
Sharp J-SH04	2000	32	122	13	74
BlackBerry 857	2000	78	116	17	140
Nokia 1100	2003	46	106	20	86
Nokia N90	2005	51	112	24	173
Sony Ericsson Xperia X10	2010	63	119	13	135
HTC One M9	2015	69.7	144.6	9.61	157
OnePlus 3	2016	74.7	152.7	7.4	158
Apple iPhone 8 plus	2017	78.1	158.4	7.5	202
Huawei Mate 10 Pro	2017	74.5	154.2	7.9	178
Samsung Galaxy S10plus	2019	74.1	157.6	7.8	175
Samsung Galaxy S21 Ultra	2021	75.6	165.1	8.9	229
Apple iPhone 13 pro max	2021	78.1	160.8	7.7	240

Figure 7: A brief overview of mobile phone/smartphone dimensions and weight from 1983-2021.

By 2010, video recording in full high definition became available in smartphones with the release of Samsung Galaxy S. Being a global broadcast, satellite and cable image distribution standard, a full HD resolution is an important benchmark for footage quality (Eve, 2014, p.2). According to Mobile Overview Report released by Scientia Mobile, the nineteen most popular handsets that cover 58 percent of the global mobile market, have at least one camera able to record a full HD resolution of 1920x1080 pixels (MOVR 2019Q3 p.5). Even complex optical zoom systems such as periscope lens[x], can be found in a smartphone, after engineers successfully scale it down in the body of Huawei P30 (Huawei).

After comparing the devices size, weight and technical sophistication, a question arises: if smartphones have been conveniently portable since 1996, and they have had video recording capabilities for the past twenty, why was there no significant record of vertical video until 2012? I would argue that the increasing size of handsets driven by smartphone manufacturers and the fortitude of social media giants to offer different user experiences become two determining forces responsible for adoption of vertical content. At the same time, the increase in the handset's width was the key factor affecting the 'comfort' of the single hand grip, subsequently prompting orientation preference (Lee & Kyung, 2019, p.61). Another study, 'Does Size Matter? Investigating the Impact of Mobile Phone Screen Size on Users' Perceived Usability, Effectiveness and Efficiency (2013), concluded:

> Mobile users that interact with a device to mainly perform information-seeking tasks, such as Internet browsing, will be more efficient if they use a device with a screen larger than 4.3 inches. The same is the case for users who mainly want to play games or watch media, as the larger screen size will lead to higher enjoyment, but with the counter-argument is that very large screens decrease the portability of the device and reduce the ability to use the phone with one hand (Raptis & Tselios, 2013, p.9).

Reports from Scientia Mobile further support these conclusions by highlighting the connection between screen size and device orientation. When watching video content, smartphone users hold their devices in portrait orientation for 82.5 percent of the time; but on tablets, the average time the device is used in portrait mode drops to 55.6 percent. Once the display diagonal is equal to or surpasses 11 inches in size,

93.9 percent of users watch videos in landscape orientation (MOVR, 2018). This relationship between the device's size and preferred user orientation is essential for vertical video consumption, since approximately 90 percent of smartphones in Asia, North America, and Europe today have a diagonal between 4.5 – 6.5 inches (MOVR, 2019Q4), (figure 8).

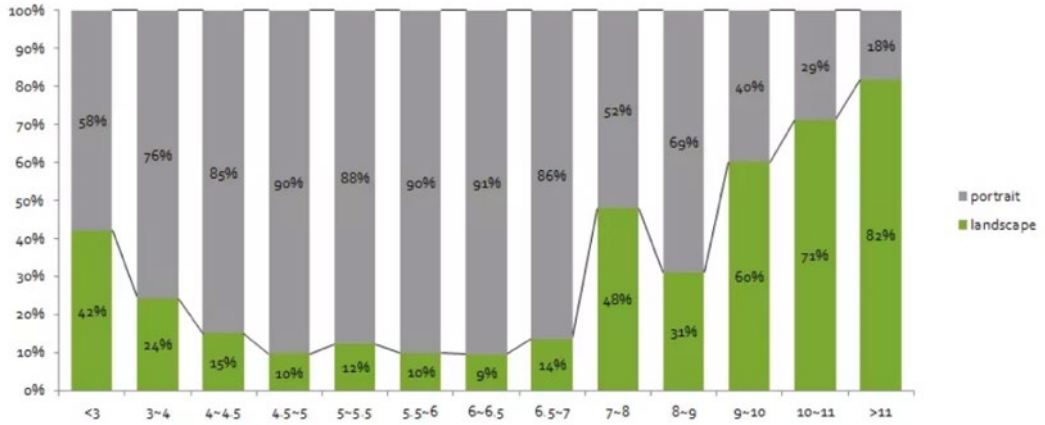

Figure 8: The influence of screen size diagonal on a smartphone and tablet orientation.

As much as the mobile screen contributed to the expansion of vertical video, these studies show that a single hand operation in upright position becomes a challenge for the user once the handset surpasses nine inches in the diagonal size. To address this problem, manufacturers like Huawei introduced 'One-Handed User Interface' that shrunk the screen by 30 percent, so that the entire content could be reached with one hand (Huawei, 2019), (figure 9). Access to 'Mini Screen View' is available in two methods depending on the user preference, and the existence of such a mode proves that a manufacturer acknowledged the benefits of a larger vertical screen from a commercial point of view and quality of user experience, but at the same time, has to address its potential limitation. With multiple companies supplying devices with advanced technical specifications, applications that deliver a better user experience become an equally important factor in the saturated mobile market.

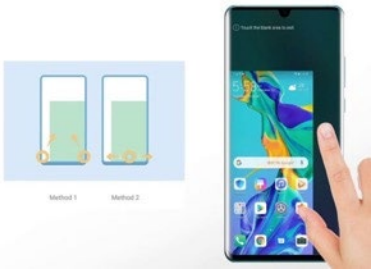

Figure 9: Huawei, 'Mini Screen View.'

One of the most recent applications that attempt to capitalise on the expansion of vertical format and, at the same time, recognise that most film and video production remained horizontal is Quibi. Launched in April 2020, Quibi targeted the short-form mobile market by offering six to ten minutes long episodes of platform exclusive content. What differentiated Quibi from other online content providers was the substantial backing of media conglomerates and a ground-breaking screening algorithm called Turnstyle. Being a veteran in the film and TV industry, the founder Jeffrey Katzenberg secured backing from major Hollywood studios, including The Walt Disney Company, NBC Universal, Sony Pictures, Warner Media, Viacom CBS, and Alibaba Group. The company advertised that 'No matter how you hold your phone, everything is framed to fit your screen' since Turnstyle will allow a seamless switch between vertical and horizontal orientation (Quibi, 2020). To achieve the uninterrupted transition, every segment on Quibi playback two video files simultaneously (vertical and horizontal version) using the audio track to keep them in sync, and ensuring there is no distraction of black letter or pillar boxing for best user experience. The playback of multiple files is part of the technical solution, while directors and cinematographers had to consider a hybrid between 16:9 and 9:16 aspect ratios while making aesthetic and compositional decisions (Hardawar, 2020). Quibi was online for only nine months, from April to December 2020, and the reasons for the lack of commercial success are still subject to speculation. What makes it interesting for this research is the fact that content produced for Quibi was created with multiple aspect ratios on a director's mind. Additionally, the budget per episode went up to $100K per minute, thus attracting 'A list' Hollywood actors, professional crew and becoming one of the most expensive vertical content available online.

In contrast to high budget productions and the immediacy commanded by social media today, another pair of authors, Marsha Berry and Max Schleser, explored expressive possibilities of mobile film making through the eyes of ethnographers, screenwriters and musicians in a collection of articles *Mobile Story Making in an Age of Smartphones* (2018). This body of work explores how smartphones influence social change and can further expand the definition of creative practices relating to the field of screen media. One of the contributors Dean Keep summarises this well in his essay 'Smartphones and Evocative Documentary Practices' (2018).

> Unlike traditional film and photographic cameras, the smartphone requires no specialised skill and provides users with a simple means of creating high-resolution digital media content. However, smartphone technology is arguably doing much more than simplifying image-making processes. Smartphones are also changing our understanding of how visual media can be used to construct, share, and screen a wide range of personal narratives (Keep, 2018, p.42).

The Internet challenged the rigidity of traditional media distribution channels, while affordability and mobility of recording technology levelled the playing field between trained professionals and amateur users. Ross and Neal acknowledge those developments in their contributing article 'Mobile Framing: Vertical Videos from User-Generated Content to Corporate Marketing' (2018) by saying: 'the slippage between amateur and professional distinctions does not mean underestimating the agency users have to work these contexts in innovative and challenging ways' (Neal & Ross, 2018, p.8).

It is clear that the progression in the evolution of tall format, like many previous remediations in the history of moving images, was steered by a combination of influences, and in 2022, multiple factors central to vertical video's integration in the vast media landscape were in full swing. Everyday Internet users, industry professionals, artists, and academics conducting research contributed to the development of the vertical format with a different point of view and their respected range of skills. But it was the expansion of technical support, endorsements of social media giants, and acceptance of global distributors of online video content such as YouTube and Netflix, that have influenced more than anything else transformation of vertical video into a more readily accepted form of online communication.

Chapter Three: Methodology - Overview

When I commenced this research with textual analysis in 2018, the vertical video could be found on all major social media platforms, which indicated it was no longer a novelty. However, I still detected echoes of the original backlash amongst my colleagues, and other media professionals, even though vertical video has experienced a significant transformation over the past several years. Inspired to learn how much vertical video has evolved and how it is affecting the language of moving images, I designed this research as a practice-led inquiry into the question:

> How do framing, filming technology, and choice of distribution platforms influence media professionals' aesthetic decisions while creating online vertical content?

The research steps included quantitative and qualitative analysis of texts, interviews with practitioners, and production of original vertical and widescreen video forms. Textual analysis encompassed vertical music videos and their widescreen counterparts existing on the YouTube VEVO channel and Netflix vertical previews available in the application search browser. I selected YouTube and Netflix as primary sources of text, since two platforms offered an abundance of professionally created vertical content aimed at the mobile audience. Detailed analysis of key texts revealed trends in contemporary filming aesthetic and editing style of vertical music videos and film trailers, while interviews with industry professionals investigated the current opinion of media professionals towards vertical content and highlighted techniques trained crews use in its production. Conclusions from the previous methodological steps informed my approach towards the practical component of the research: making the thesis artefacts.

While producing two original vertical and two widescreen single-take music videos, I reflected on interview findings and empirically tested expressive possibilities of vertical video against a widescreen aspect ratio when filming single versus multiple subjects. At the same time, I looked into affordances between smartphone usage as a primary recording device instead of DSLR/SLR and larger camera bodies in the professional workflow, focusing on footage aesthetics prescribed by the device technical capabilities and form-factor. Lastly, acknowledging that most newly generated online video content is still recorded in one of the widescreen aspect ratios, I edited the fifth and sixth artefact, a vertical film and 1:2 film trailer, from cinematic 16:9 widescreen footage. In this last methodological step, I examine the editor's rationale that accompanies material re-purposing thought processes and methods applied in nonlinear editing software. A combination of methodologies gave me a robust framework to investigate multiple facets of the principal thesis question and examine the vertical frame's technical and aesthetic parameters during its production in the professional workflow.

Chapter Three: Methodology - Finding the Source of Relevant Texts

In 2021, the rapid evolution of vertical video was visible not only in the quality and variety of produced content, but also in the development of unique aesthetics on major online distribution platforms such as YouTube VEVO channel. Being the largest video hosting platform, and the second-largest search engine globally, YouTube host thousands of channels on the platform, and despite the profusion of other online and legacy media music channels, only a few have such strong backing from some of the major conglomerates in the music industry like the YouTube VEVO channel. YouTube VEVO is a video hosting service jointly owned by Universal Music Group (UMG), Google, Sony Music Entertainment (SME), and Abu Dhabi Media. Founded in 2009, VEVO became popular as a content aggregator since the platform can market the videos under a unified brand. In 2020, VEVO reported 26 billion views every month, with 60 percent of those coming from a mobile device. Additionally, out of the ten most popular videos on YouTube, VEVO channel lays claim to six of them (VEVO, 2020).

Since the supplementary thesis questions probed creative and technical decisions related to the footage being re-used in a non-original aspect ratio, I looked for an online source of the material that fit those criteria. Similar to YouTube, Netflix was a prolific source of professionally created vertical content.

Vertical previews for TV shows and movies featured in the application browser were all edited from the original widescreen material, and by being framed in a 1:2 aspect ratio, they represent the tallest vertical frame featuring a narrative storytelling sequence that I could find online. Furthermore, in the increasingly saturated market of OTT content providers, Netflix holds the largest share. According to the Motion Picture Association's 2019 Theme Report, at the end of 2019, Netflix had 167 million subscribers, servicing 19 percent of the global digital streaming market (MPA, 2019, p.34). In the years after, that number continued to grow as MPA's theme 2021 report shows (MPA, 2021, p.5). Although the company reported the first-ever quarter without adding new customers in June 2022, Netflix is still the largest OTT provider in the world.

In searching for relevant samples, I initially looked at several other online distribution channels: Facebook, Instagram, Snapchat, TikTok, Twitter, and LinkedIn, before limiting the scope to YouTube VEVO and Netflix previews produced during 2017, 2018, and 2019. Being the largest social media platform, comprised of almost three billion diverse global users, Facebook was my starting point. Even though preliminary research discovered vertical videos hosted on Facebook had a more extensive engagement and completion rate on average, the interactivity of Facebook's multiple feeds made the ranking of content by popularity a challenge. I found similar issues on Instagram, where the existence of text in aggregated feeds complicated structured search. Moreover, Facebook or Instagram do not have any channels dedicated to vertical video or professional content only, which further obscured the selection of appropriate texts.

Samples found on other popular social media outlets such as Snapchat and TikTok also didn't fit the research criteria. In contrast to the broad demographics found on Facebook, Snapchat and TikTok mainly targeted younger audiences and encouraged user-generated, not professional content. Finally, like Facebook and Instagram, Twitter and LinkedIn do not have channels featuring only vertical videos. Being aware of the dynamics related to the rapid evolution of technology in the mobile media realm, I continued to scout the field for new developments even after identifying the online channels most relevant for the research. The time invested in the continuous survey proved beneficial in April 2020 with a surprise launch of Quibi, an application wanting to bridge the gap between vertical and widescreen content I mentioned briefly in chapter two. Coming to an already busy market of mobile media, Quibi developed its strategy around short-form news and entertainment videos, or 'quick bites', running ten minutes or less in length. It attracted top Hollywood stars for its shows and high-profile investors, including legacy media companies like Disney, Comcast's NBC Universal, and AT&T's Warner Media (CNBC, 2020).

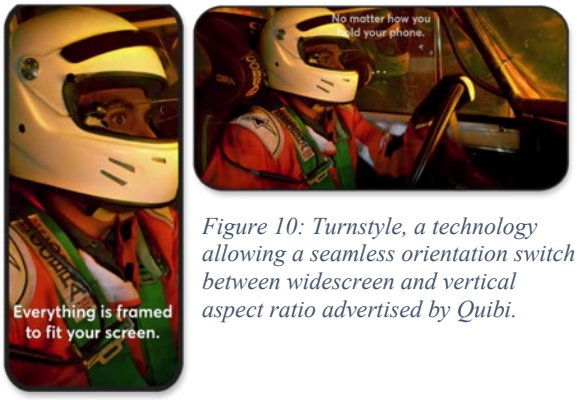

Figure 10: Turnstyle, a technology allowing a seamless orientation switch between widescreen and vertical aspect ratio advertised by Quibi.

What made videos found on Quibi highly relevant for this research wasn't only the quality of produced content but the fact that this was the first application to give equal prominence to horizontal and vertical screen orientation. In the back end of Quibi, a newly developed technology named Turnstyle allowed a user to seamlessly switch between screen orientations by streaming two files simultaneously, (figure 10).

Although highly intriguing, the complexity of Quibi's back-end streaming systems and pre-production, filming, and post-production workflow open multiple side questions that fell outside of the scope of this research. Furthermore, after only six months of operation, Quibi went off-line in December 2020, and today it is not possible to watch any of the Quibi shows in their original form. Due to its importance as a prime example of professionally created cinematic and narrative vertical content, I included interview questions related to Quibi focusing on footage re-purposing aspects, but I excluded Quibi from the textual analysis in exchange for examples from more enduring platforms. To summarise, out of all listed distribution platforms, the

abundance of texts, availability of organised feeds, and unique footage characteristics made YouTube VEVO vertical channel and Netflix browser key data sources for textual analysis.

Chapter Three: Methodology – Selecting Relevant Texts

While reviewing hundreds of music videos in the initial stage of the study, I looked for titles with an official widescreen and vertical music video with a minimum of one million or more views. The search criteria rendered a list of 94 labels with an official widescreen and vertical version and multimillion viewing numbers. Still, this pool of samples also encompassed multiple user-generated music videos. The aesthetics and technical solutions in some of the amateur videos were stimulating, and the viewing numbers were a testament to the power of amateurs on platforms with an immense reach like YouTube. But because I was researching professionally created vertical content, in the second pass, I refined the criteria. This time I looked for critically acclaimed, award-winning artists, holders of awards such as Grammy, Billboard Music, or The BRIT accolade, i.e. performers who, without a doubt, can afford a professional crew in every step of music video production. I also increased the bar to a minimum of one million views for the official vertical music video, and the new criteria presented a catalogue of 11 titles. The listed tracks best represent contrasts between official widescreen and vertical versions of their respective music videos, except *Billie Eilish - I Don't Wanna Be You Anymore*. That is the one track on this list released as vertical video only, with no official widescreen equivalent, but the unorthodox colouring of pillar boxes exhibited when the official vertical track is watched on widescreen canvas, (figure 11), brands this video intriguing for analysis and comparison.

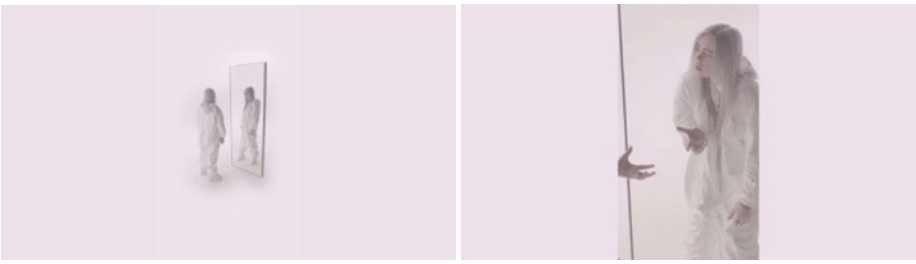

Figure 11: Vertical music video I Don't Want to be You Anymore (2017), projected on widescreen canvas with pillar boxing matching the colour pallet of production design.

This is a list of key samples selected from YouTube VEVO. For readers ' convenience, I included hyperlinks to originally posted widescreen videos and a separate link to vertical versions.

YouTube Sample #1 Mark Ronson - Nothing Breaks Like a Heart ft. Miley Cyrus / Vertical Version
YouTube Sample #2 Bebe Rexha - Last Hurrah / Vertical Version
YouTube Sample #3 Billie Eilish - You Should See Me in a Crown / Vertical Version
YouTube Sample #4 Taylor Swift - Delicate / Vertical Version
YouTube Sample #5 Billie Eilish – I Don't Wanna Be You Anymore, no Widescreen Version
YouTube Sample #6 Camila Cabello - Havana ft. Young Thug / Vertical Version
YouTube Sample #7 Katy Perry - Harleys In Hawaii / Vertical Version
YouTube Sample #8 Nicki Minaj - Chun-Li / Vertical Version
YouTube Sample #9 Sam Smith - How Do You Sleep? / Vertical Version
YouTube Sample #10 Sam Smith - Too Good at Goodbyes / Vertical Version
YouTube Sample #11 Halsey - Without Me / Vertical Version

During the preliminary screenings while looking for more relevant examples of professionally created vertical music videos, I expanded the investigation to two of YouTube's other major musical channels, Sony Music Entertainment and Warner Music Group. The survey identified an additional 46 samples that fit the refined criteria where tracks had an official widescreen and vertical music video and at least one million views. Still, none of the texts exhibited unique aesthetic characteristics I observed on the YouTube VEVO channel, prompting me to concentrate on the already enlisted platform. Equally important, a deeper investigation into YouTube VEVO channel business arrangements revealed VEVO

signed a licensing exchange agreement with Warner Music Group in August 2016 (Warner, 2016). Before signing the deal with Warner Music Group, the VEVO channel already owned the largest catalogue of on-line music videos, still, after including the last independent record label conglomerate, it became an unrivalled hub for A-list names in the global music industry. Therefore, I decided to focus my research on the YouTube VEVO channel as the leading platform for professionally created vertical music videos. I will unpack the discussion on principal aesthetic differences of key analysed texts in chapter four.

Owning the equivalent complements of YouTube in the domain of global streaming services, the library of the world's #1 provider of OTT content, Netflix, offered a myriad of samples in assorted genres, but categorization of texts according to their popularity was not possible in this case. OTT providers are notoriously secretive about their streaming numbers, and as Amanda Lotz has noted, 'subscriber-supported platforms have a little economic incentive to publicise their audience data' (Lotz, 2017, p.14). As the lack of platform transparency prohibited me from verifying the streaming data, I opted for diversity in the sample instead of measuring levels of popularity. To provide gender and age range diversity in the viewing preference, I sampled film and TV series vertical trailers from five individual accounts used by the same person for at least one year. Profiles belonged to one girl, two adult females, and two adult males. The account holders ranged from seven to 71, with an average age of 36.2 (f7y, f44y, f70y, m18y, m42y). In the search for relevant samples, I initially screened approximately 700 trailers, looking for an average of ten per genre from each account. Downloaded material included trailers for the following categories: Documentary film, Sport, Reality-TV, Comedy, Drama, Romance, Adventure, Fantasy, News, Talk-Show, Horror, Game-Show, Musical, and Animation. From the original volume of data, I selected eleven texts in various genres for detailed analysis, looking for the distinctive aesthetic qualities of the footage and editing styles. Similar to the methodology applied in the study of music videos, I juxtaposed vertical previews with their original widescreen counterparts to highlight the differences in dominant two and three-dimensional compositional elements between them. A list of all samples included in the survey is available in appendix #2. Here is the summary of the key examples selected for detailed analysis. As in the case of key samples the from VEVO channel, I included the hyperlinks to the original widescreen version of selected trailers except for vertical previews. Netflix vertical previews originated on personal feeds, so there was no direct link for the media. Looking to comply with the copyright rules of both platforms, I uploaded all research material as 'unlisted' on YouTube, so it cannot be discovered in user Internet search, but a reader can access the files via links provided in this document.

Netflix Sample #1 Formula 1 - Drive to survive S1 / Vertical Version
Netflix Sample #2 Glow Up: Britain's Next Make-Up Star S1 / Vertical Version
Netflix Sample #3 Inside Bill's Brain: Decoding Bill Gates / Vertical Version
Netflix Sample #4 Marriage Story (2019) / Vertical Version
Netflix Sample #5 Mowgli: Legend of the Jungle (2018) / Vertical Version
Netflix Sample #6 Orange Is the New Black S1 / Vertical Version
Netflix Sample #7 Patriot Act (2018) / Vertical Version
Netflix Sample #8 Stranger Things S1 / Vertical Version
Netflix Sample #9 Sugar Rush S1 / Vertical Version
Netflix Sample #10 The Ballad of Buster Scruggs (2018) / Vertical Version
Netflix Sample #11 White Fang (2018) / Vertical Version

Depending on the source of material and available playback options, I combined different programs to download files for study. Being a video sharing platform, YouTube commonly allows direct downloads, introducing protections only if a person uploading specifically asks for one. I used paid software, 'Media Human YouTube Downloader', to capture material from the VEVO channel because it features a simple user-friendly front end interface, capable of converting multiple video links, and the entire playlists into $H264^{x}$ compressed files. This was my first choice as a straightforward solution to acquire high-quality video footage packed in small files convenient for analysis and efficient off-line storage. In contrast, I had to use two additional steps to capture footage from Netflix accounts on a hard drive as the platform doesn't allow downloads of vertical previews on a personal computer. Video footage

from the Netflix desktop application was recorded with Mac OS proprietary Quick Time Screen Capture tool in the first pass, while the recording of audio required a second pass using the same feature. Obtaining video samples from Quibi was even more complex, as 'screen capture protection' embedded in application code prevented recording screen shots or video clips during playback. Without third party or OS built-in software able to circumvent the protection, I had to record a video from a smartphone screen with an external camera in 16:9 full HD resolution. The audio in Quibi application did not have any protections, so I was able to capture it with built-in Apple OS feature during the second pass. Once I had the data on the hard drive, I synchronised video and audio files in Adobe Premiere Pro[x], and re-exported samples as a single full HD file for easier data management and backup.

Since my proposed methodology involved classifying each separate shot, I used the automated feature 'Scene Edit Detection' built into Premiere Pro to identify edit points in the footage and create individual sub-clips. Once the automated system identified edit points in the source files, I visually verified that the algorithms prediction was correct, and in approximately ten percent of cases made manual adjustments if fast camera movements or lack of image focus misled the AI. The following step included classification of shots by scale, from vista to extreme close-up, and designation of separate timeline tracks for each group. From that point, a simple 'lasso' selection of all clips on a separate track was providing an exact number of shots in each group in Premiere Pro-info panel, (figure 12).
As I worked on the sample analysis, the 'Export Frame' feature built into the Premiere program monitor proved to be a convenient tool, able to quickly export screenshots needed to illustrate research points.

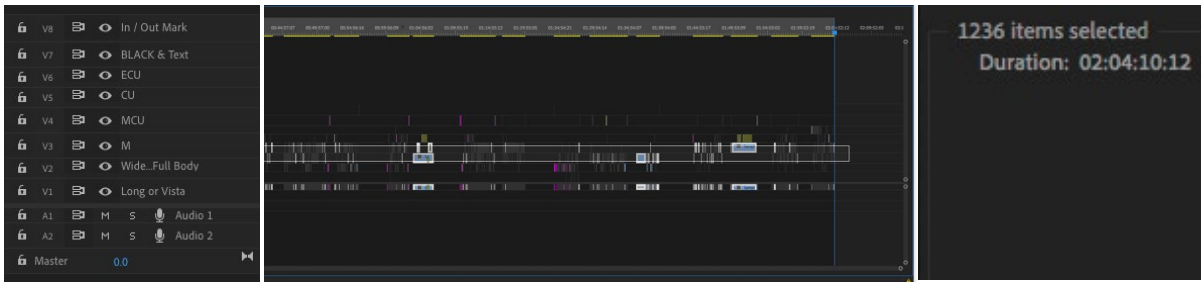

Figure 12: Footage classification by shot scale. Lasso tool on an Adobe Premiere Pro timeline and info panel showing a selected number of items.

Although automated features of Adobe Premiere Pro needed some manual adjustments, the described workflow substantially streamlined the process of footage classification, and improved my ability to identify trends in the analysed texts. Finally, to comply with YouTube and Netflix copyright regulations, all gathered files, including derivatives generated during the research, stayed on a password-protected hard drive accessible to principal researcher only, and won't be released to the public. Once the research is concluded, all data will be deleted except key samples and thesis artefacts that I will deliver to the Massey University library archive on a separate hard drive.

Chapter Three: Methodology – Text Analysis

Since the preliminary survey of vertical samples already revealed an almost complete absence of vista shots in favour of an increased number of medium compositions, I started the analysis with a quantitative step by looking for the dominance of particular framing in the footage. According to AMA, 'the pre-eminence of a few parameters in the footage dictates the aesthetic style of the piece' (Zettl, 2016, p.13). These are the parameters I scrutinised while looking for trends:

- the total length of the video and number of shots
- classification of footage by shot scale
- a number of shots having clear black screen and text only
- the average length of shots between two edit points

By categorising these listed parameters, I obtained reliable data that described the editing rhythm and tempo as well as some information on the methods that crews used during production. This was

important when juxtaposing widescreen and vertical music videos of the same track, particularly in samples where the first version featured more than a hundred cuts and the other just a few. Furthermore, during the analysis of Netflix vertical previews, I extended the survey of the shot scale by counting the number of closed captions title cards, the average number of words used in them, and the total number of characters visible on the screen for the duration of the trailer. To accurately and efficiently count the number of characters used for closed captions, I uploaded the JPEG screenshots from Adobe Premiere Pro timeline into Google Drive and used its proprietary optical character recognition (OCR) feature. Unlike cinema, where the position of closed captions is standardised, I notice that the screen placement, the number of words and characters, and the font size are not consistent between cuts. As the analysis in chapter four will demonstrate, the variety in closed caption title cards created an additional visual sensation on the screen and significantly affected the audience judgement of principal footage.

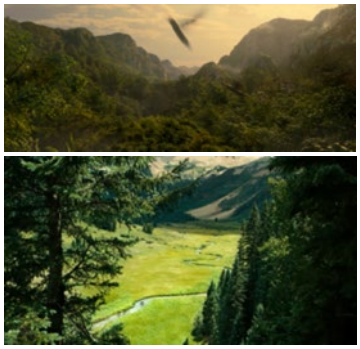

Figure 13: V shot. Mowgli (2018), and The Ballad of Buster Scruggs (2018).

For an easier understanding of shot scale and my analysis, below I include an overview of the traditional definitions with illustrations. To cover the most common television and cinema aspect ratios, I selected examples in the range from 4:3 to 2.39:1.

Shot size or camera-subject distance is a standard measure used in film theory and cinematography to describe how much of a human figure is present in a shot (Salt, 2006, p.423). Like any other demarcation in the following list, the exact look and size of the frame depends largely on the principal subject and the rest of the context in the diegetic space.

Vista (V) – As a rule of the thumb, a vista shot is any shot that encompasses the entire scene in which it is hard to differentiate a single human body from the rest of the scenery, (figure 13).

Figure 14: W Shot, Katy Perry – Harley's in Hawaii.

The wide shot (W) includes a broad section of the location, but unlike the vista shot, a human body is distinguishable in wide composition. As instruction for camera operators, a director can have more specific guidelines depending on the aesthetics and script. For example, to create a large leading room for motorcycles and emphasise the feeling of open space on the road, the cinematographer selected a frame where the key subject can fit 4-5 times its size on the horizontal axis and 2-3 times on the vertical as you can see in the example of wide shot in figure 14.

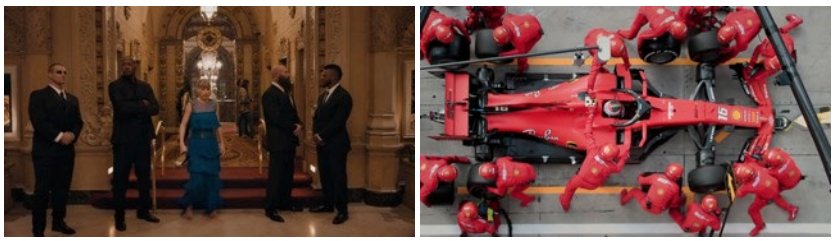

Figure 15: Full body shot from Taylor Swift – Delicate (2017), and Formula 1- Drive to Survive (2019).

A full body shot (F), indicates that we see the character from head to toe or the entire object of interest comparable to the size of an adult human (Brown, 2016, p.62). For example, a shot of a car would include the whole vehicle with some available space around it, (figure 15).

Figure 16: M Shot, Luther (2010).

The next in the sequence is a medium (M) shot, (figure 16). As in the case of wide, the boundaries of a medium shot are relative to the size of a principal subject. Assuming a human protagonist is a point of interest, once the camera moves closer to the subject, the audience can read more of the talent's facial expressions, details of how they are dressed, and other intricate mise-en-scene elements less prominent in a wide composition. In general, a medium shot of a person starts from the waist up and allows at least ten percent of headroom (Marscelli, 2005, p.24).

A definition given in *5 C's of Cinematography* by Joseph V. Marscelli is an excellent example of shot scale inconsistency, and space left for context dependent interpretation. For example, a medium shot in dramas would commonly include a body of a person from the waist up, as you can see in the figure 17, but that is not the case for Hollywood Westerns. Due to the necessity of including a low hanging gun, a medium shot in Westerns was commonly 15 to 25 percent wider, having a bottom edge around the protagonist's knees.

Figure 17: Variation of medium shot between aspect ratio and genre. 4:3 Europe drama vs 2.35:1 Western movie. From left to right: Persona (1966), Breathless (1960), Once Upon a Time in the West (1968).

Suppose I consider the properties of diverse widescreen standards and the difference in the size of the X and Y axis between them. In that case, it is understandable why there is no one unifying definition for any frame in the shot scale.

By moving the camera to a Medium Close-Up (MCU) position, encompassing a person from mid-chest with clearly perceivable facial expressions, the director/cinematographer often suggests the increased importance of spoken words. At the same time, some context of the scene is still present in the MCU composition as you can see in the figure 18. As the camera continues to approach the subject, tighter framing further exemplifies the emotion or details on the object of interest, (figure 19). Once again, the exact boundaries of a Close Up Shot (CU) may be detailed in the script according to image size or object of interest, or alternatively left to the discretion of the director or cameraperson (Marscelli, 2005, p.25). Finally, an Extreme Close-Up (ECU) singles out a small element of the principal subject of interest and, if used seldom, can produce a strong accent in the storyline, (figure 20).

Figure 18: MCU, Bebe Rexna Last Hurrah (2019).

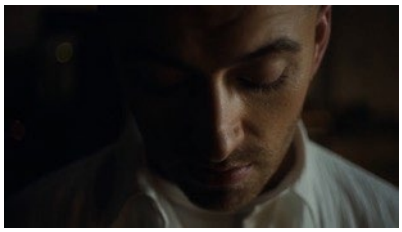

Figure 19: CU Sam Smith - To Good at Goodbyes (2017).

Figure 20: ECU - The Hobbit: The Desolation of Smaug (2013).

Once I quantified the technical parameters of the footage, I continued to investigate the principal aesthetic characteristics of the text. These are the two-dimensional factors I looked at:

A) magnetism and asymmetry of the frame/edge pull,
B) predominance of a specific directional vector,
C) objects figure/ground relationship,

as well as prevailing three-dimensional structuring elements such as:

D) presence or absence of distinct foreground, middle and background sections in the shot responsible for Z-axis articulation,
E) presence or absence of secondary framing,
F) and classification of shots where the main protagonist uses a direct address method.

The following sections will provide more details of these listed categories.

A) Magnetism and asymmetry of the frame, a.k.a. edge pull

The main subject's proximity or distance from the edge of the frame is one of the decisive elements of mise-en-scene, as it dramatically affects the overall perception of the shot. As Arnheim observed in *Art and Visual Perception*:

> No object is perceived as unique or isolated. Seeing something involves assigning it a place in the whole: a location in space, a score on the scale of size or brightness or distance. What a person or animal perceives is not only an arrangement of objects, of colours and shapes, of movements and sizes. It is, perhaps first of all, an interplay of directed tensions. Because they have magnitude and direction, these tensions can be described as psychological 'forces.' At the centre all the forces balance one another, and therefore the central position makes for rest (Arnheim, 1974, p.11).

As you will see from multiple examples in chapter four, the central position is the most stable in widescreen and vertical shots because significant headroom or free space to the left and right would suggest unbalanced composition, which hints a protagonist's emotions or state of mind. However, the difference in the length of the horizontal axis still renders two very distinct compositions, even if they are built from similar components. Likewise, asymmetry of the frame is a phenomenon that describes the ambiguity of human perception concerning the dominance of visual elements and their position within the composition. According to Zettl, the right and left sides of the video, film, and computer screen are unequal in visual prominence because the right side commands more attention than the left. 'This phenomenon is even more prominent in the Western world, where the text is written from left to right, making our minds pre-wired to scan the image in a similar order' (Arnheim, 1982, p.37).

Figure 21: Strong edge pull. Die Hard (1988).

In combination, magnetism and asymmetry of the frame produce a phenomenon described in AMA as 'edge pull' (Zettl, 2016, p.143). The edge pull of the frame is a perceptual characteristic of the composition, where proximity to the edge suggests a subject's connection to it. The bottom edge of the frame is commonly perceived as ground, but if any of the other three edges exhibit a strong pull, the ground perception can shift. As you can see from the example in figure 21, the right screen edge cuts the protagonist's body, signifying their connection and strongly suggesting a presence of the off-screen space behind it. In contrast, another protagonist has free space before the left screen edge and stronger luminance levels on his side of the frame. Furthermore, a Dutch camera angle produces the impression one protagonist occupies 'higher ground', suggesting a balance of power in a story and putting more weight on the character in the corner. As a result of such framing, blocking, and mass distribution, the perception of the most stable edge of the frame, shifts from traditional bottom towards the right screen edge in this composition. In the context of the *Die Hard* (1988) story, John McClane, is offering the weapon to Hans, a character previously established as an antagonist, but the blocking infers distance between them and John's desire to be in the 'off-screen space,' away from the villain.

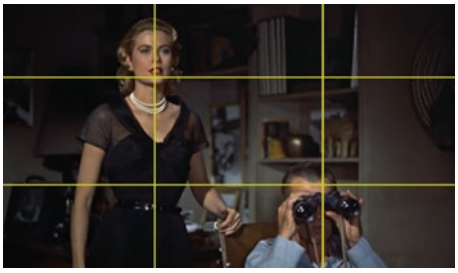

Figure 22: The rule of third on the screen shot from Hitchcock's Rear Window (1954).

To control magnetism and asymmetry of the frame, visual content creators resort to 'the rule of thirds'. The rule of thirds proposes that the approximate starting point for any compositional grouping or major points of interest in the scene should be placed in one of four intersections formed by grid lines, see figure 22 (Brown, 2016). Similarly, like shot scale, the rule of thirds is hard to define precisely because shot balance is influenced by multiple mise-en-scene elements and the image aspect ratio. Despite some ambivalence in the definition, this convention is a simple and practical guideline for assembling a balanced

composition in any widescreen aspect ratio. However, once translated to vertical arrangement, a short horizontal axis of an image brings this convention into question. I will elaborate more on this observation in chapter seven through thesis items #5 and #6, a vertical film and trailer edited from original widescreen material.

While scrutinizing the samples' magnetism and asymmetry of the frame, I categorised their presence on the following facts. If a principal subject was close to the corners or edge of the frame in 50 percent of shots, I labelled the sample with 'strong magnetism'. In case the rule of thirds was conformed to for more than 50 percent of running time, I categorised a sample with detectable magnetism. Finally, if a subject was placed in the centre of the composition or very close to it for more than 50 percent of the time, it marked a weak magnetism in the sample.

B) Predominance of a directional vector

AMA argues that directional forces are one of the most potent forces operating within the screen. There are three principal types: graphic, index, and motion vectors (Zettl, 2016, p.127). A stationary element creates a graphic vector that guides our eyes in a specific direction, for example, a shape of the river or a converging lines on the road, (figure 23). The most common example of an index vector is the direction of a protagonist's point of view, (figure 24), while the motion vector gets induced by an object moving in a specific path, (figure 25). Graphic, index and motion vector can be present in the same composition, but out of the listed three, the motion vector is the most dominant in the group for the reason that it will attract the spectator's attention with great intensity. In combination with mise-en-scene components, these screen forces lead spectators to gaze from one point to another.

Figure 23: Graphic vector induced by the shape of the river and mountain.

Figure 24: Index vector produced by protagonist gaze.

Figure 25: Motion vector prompted by movement of the cars, dominate over index vector induced by road markings.

While searching for the dominant vectors, it is important to note, the primary motion is one coming from mise-en-scene, secondary is produced by camera movements, while tertiary motion originates as a result of the footage sequencing in the editing (Zettl, 2016, p.294). Additionally, the acuity of all directional vectors is accentuated if the sequence is built on the conventions of continuity editing. As Tim Smith pointed out in *The Attentional Theory of Cinematic Continuity*:

> the continuity editing rules use natural attentional cues such as off screen sounds, conversational turns, motion, gaze cues, and pointing gestures to trigger attentional shifts across cuts. The combination of attentional cues pre-cut, and matching minimal expectations post-cut allow viewer cognition to precede seamlessly from shot to shot, scene to scene, sequence to sequence, and across the entire narrative (Smith, 2011, p.5).

In the case of single-take videos where the camera stays stationary most of the time, and there is no editing sequencing, the main protagonist blocking predisposes the directional vectors and acts as the main driver of the audience's attentional synchrony[x]. I categorised directional vectors as strong if the footage is dominated by a motion vector produced by the protagonist object or camera movement in 70 percent of shots. Detectable, if index vectors created by the protagonist's eye-line were the most prominent screen force leading the audience gaze with minimal or no camera movement in 70 percent of shots. Or weak if the prevailing vectors were graphic, i.e., influenced by static objects and scenery in 70 percent of shots.

C) Objects figure/ground relationship

A human brain tends to organise a scene into figures that lie in front of a background, and in doing so, we perceive the ground as being more stable than the figures, (figure 26). Contrary to the static and fixed figure/ground relationship in a painting or still photograph, the on-screen relationship changes (as it does in real life) with the camera's point of view (Zettl, 2016, p.120). According to Rudolf Arnheim, the frame as we know it today developed during the Renaissance from the facade-like construction of lintels and pilasters that surrounded the altarpieces.

> This world came to be conceived as boundless-not only in depth, but also laterally-so that the edges of the image designated the end of the composition, but not the end of represented space. The frame was thought of as a window, through which the observer peeped into an outer world, confined by the opening of the peephole but unbounded in itself (Arnheim, 1974, p.239).

Figure 26: Bottom edge is commonly perceived as ground, and the most stable one out of four.

Similarly to paintings, the distribution of mass within the video composition, in combination with a choice of camera angle and the amount of headroom, provides clues about the off-screen space and determines which screen edge will be perceived as dominant by the audience. In this instance, I categorised objects' figure/ground relationship as strong if there is a limited amount of headroom and the figure (or figures) are stable in relation to the bottom edge of the screen (figure 27).

Figure 27: Strong figure/ground relationship.

Figure 28: Detectable figure/ground relationship with a strong hint of the top off-screen space.

Figure 29: Weak figure/ground relationship. Stability of the bottom edge is challenged by motion vector.

Detectable if figures are detached from the ground but the lower edge of the frame is still perceived as 'stable off-screen space', (figure 28), or weak if the stability of the ground is challenged by index or motion vectors, as you can see from figure 29.

D) Presence or absence of distinct foreground, middle and background sections in the shot responsible for Z-axis articulation

Articulation of the Z-axis in 2D space is a complex topic in itself, and for that reason, in this research, I focused on the dominance of movement along the Z-axis, in combination with depth of field as a primary method of attention direction. Blocking the subjects is not a unique aspect of Z-axis articulation, as the size and optical characteristics of a lens, combined with the choice of aperture, greatly influence the perception of depth in the composition. Moreover, optical lens characteristics are also influenced by its diameter and the dimensions of the processing sensor it is designed to cover. This is where the choice of recording device will profoundly affect the operator's mobility, the aesthetic characteristics of recorded material, and the post-production possibilities. I will further discuss how the choice between a smartphone and a large camera body affects the visual features of the footage in chapter six, through the example of single take music videos I filmed with a smartphone as a primary recording device.

In the analysed text from YouTube VEVO and Netflix, I classified the footage as having strong Z-Axis articulation if in a sequence of the shots, seven out of ten on average, the main protagonist or principal item of interest moves mostly along the depth of the shot, and compositions rely on deep focus to tell a story, (figure 30). If a similar arrangement is observed but dominates less than 50 percent of shots, Z-axis articulation is characterised as detectable, (figure 31). Finally, if an accent in the storyline is made through the use of shallow depth of field on a single focal plane (no focus pull) in less than 30 percent of shots, I mark Z-axis articulation as weak, (figure 32).

Figure 30: Strong Z-Axis articulation produced by deep focus and subjects moving along the depth of the shot.

Figure 31: Detectable Z-axis articulation.

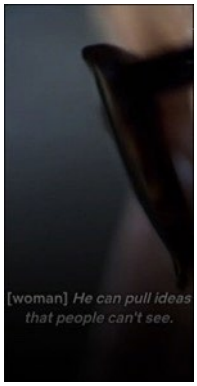

Figure 32: Weak Z-axis articulation caused by the shallow depth of field.

E) Presence or absence of secondary framing and volume duality

Figure 33: Examples of secondary framing in the composition.

As another hallmark of elaborate mise-en-scene, a secondary framing commonly suggests subtext in the story. Depending on the script, connotation of secondary framing might include the protagonist's state of mind, relationship to other characters, or ability to 'freely' select the path in the story to name a few. A good example of secondary framing are door frames and long hallways, (figure 33). Additionally, the distribution of positive and negative volumes within the arrangement accentuates or diminishes the presence of secondary framing and brings further connotation into the composition.

Buildings and similar structures are examples of positive volume, while an empty sky or space covered in uniform colour is a good illustration of negative volume in the composition, (figure 34).

The interplay between positive and negative volumes is called volume duality (Zettl, 2016, p.190). I categorised the presence of secondary framing as strong if the central character or object that drives the story was situated within mise-en-scene elements that create positive volume in 70 percent of shots, detectable if used in less than 50 percent, and weak if used occasionally as a story driving element in less than a 30 percent of the shots.

Figure 34: Example of a large negative space induced by a single colour, against multiple positive volumes of pillars and mirrors that surrounds the protagonist.

F) Classification of shots where the main protagonist uses a direct address method

Being the strongest representation of the performance convention known as the 'fourth wall breach,' the direct address method is an efficient way for an on-screen talent to establish an emotional connection with the audience quickly. AMA describes the direct address method as a 'composition in which performers speak directly to us, the viewers', (figure 35) (Zettl, 2016, p.236). It is also important to note that the perception of the direct address is significantly affected by the screen size. A screen able to reproduce a human image considerably larger than the actual size offers more details but also brings distinct connotative meaning into diegetic space, i.e., due to a significant difference in projection size, TV and mobile audiences will perceive the CU of a human face as more intimate than a theatrical audience will. In this research, the direct address method is labelled as strong if the protagonist looks or gestures at the lens in more than 70 percent of shots, detectable if present in 50 percent, and weak if used in less than 30 percent of all shots.

Figure 35: Examples of direct address method in vertical and widescreen music video.

Chapter Three: Methodology – Identifying Interview Participants

My working paradigm for semi-structured interviews was to explore questions related to the contemporary professional production of vertical content, and inquire into the level of specialised support available for its seamless integration into high-budget workflows. Instead of looking to collect a more extensive data set, I opted for the variety of age, specialization, and participants' geographic location. But, without a unifying definition of the term 'media professional', I spent a significant amount of time refining criteria to qualify participants. Because this research often reflects what was traditional and the new conventions in the language of moving images, I set up the following prerequisites for the potential participants. All candidates needed to have at least an undergraduate degree in the field of media, journalism, communication, or film. Such a degree would secure their understanding of essential terminology during the interviews. Participants' portfolios had to include at least twenty pieces of individually produced content, ten projects created in a team having a minimum of five crew members, and at least five years of experience in a production company or institution with ten or more full-time employees. The requirement for a minimum of five years' experience was vital to make sure that interview participants were already in the role before a significant expansion of vertical content in 2017. Lastly, the size of the team and delegations of tasks reflects the size of the budget, as the ability to hire a crew for specialised roles is one of the hallmarks of professional production.

To reach out to potential participants, I used the snowball sampling technique by spreading the word in personal and colleagues' professional networks. Jerry Willis in *Foundations of Qualitative Research: Interpretive and Critical Approaches* (2007) stresses that 'the worldview within which you are conducting research plays a defining role in how you prepared for the interview, who you choose an interviewee, what questions you ask, how you structure the interview, and how you interpret the data' (Willis, 2007, p.194). I concur with the author, so I took into account that two participants are my former colleagues, one academic and the other professional, and our personal connection would make them feel more comfortable during the conversation. I was also wary of the jargon and colloquial language commonly used by crew that worked together frequently. For that reason, I actively steered away from any segment of professional workflow related to our mutual projects, and questions that weren't related to the topic of this research. Furthermore, the semi-structured interview form also provides a guide template for the researcher that allows for a more conversational approach to the exchange. Thomas R. Lindlof and Bryan C. Taylor (2002) call this a 'narrative' interview that results in a 'dyadic interaction' (Robinson, 2012, p.179). This kind of interaction was appropriate as I was looking for participants in my former professional network, but equally important, as Lyn H. Lofland and John Lofland suggest, I

could draw upon personal knowledge and background to identify relevant participants in a larger familiar group (Lofland, 1995).
Lastly, in order to provide transparency while simultaneously looking for guidance of senior academic faculty through the interview process, I submitted a high risk ethics application to the Massey University Human Ethics Committee, #NOR 20/67, and got approval for my methodology and questions. A complete list of questions is available in appendix #3.

Of the six final participants, the first two work in the United States, one on the East Coast in New York and the other in the film industry in Los Angeles. The third interviewee currently works in Spain, the fourth in Australia and the fifth and sixth in New Zealand. Occupations and professional focus of the group varied from multitasking freelancers to media studies university professors. Interviewees ranged in age from 25 to 55, with an average age of 41, and a portfolio spanning between five to 30 years of continuous employment, with an average of 18.3 years of experience. All participants were familiar with television production; four have mixed portfolios spanning between film, TV and OTT streaming services, while three also have backgrounds in academia. Five subjects agreed on the use of their real names, while one decided to participate under the pseudonym. I conducted two interviews in person and the remaining four via Skype. All audio recordings were approved by the respective participants, and the researcher transcribed the audio recordings for analysis. Here is a brief overview of participants biographies:

Anna: age 25, five years in the industry.
Location: Wellington, New Zealand.
Current position, Head of Content at 'Re': TVNZ.
In her own words, same as many others from the millennial generation, she started as a one-person multitasker while she refined her specific TV production skills and eventually advanced to a management role. She now coordinates a team of seven reporters that are all versed as multitasking video content creators.

Elsie: age 33, 14 years in the industry.
Location: Auckland, New Zealand.
Current position: Freelance editor/first assistant editor.
Elsie is highly proficient in set operations as she often worked as a digital image technician and video editor on set. Unlike other participants, Elsie was never interested in multitasking roles and always focused only on the post-production aspect of media creation workflow during her entire career.

Max: age 41, 18 years in the industry.
Location: Melbourne, Australia.
Current position: Associate Professor of Film and Television at Swinburne University of Technology
As a freelance filmmaker, Max uses a smartphone as a primary recording device. He is a strong advocate of mobile filmmaking and one of the pioneers of mobile storytelling. His early cell phone films are now more than twenty years old.

Roi: age 45, 18 years in the industry.
Location: Vigo, Spain.
Current position: Freelance filmmaker, theatre performer, concept artist and designer.
Roi switched his focus from indie filmmaking to theatre performances and augmented reality experiences during the last five years. He is proficient in directing and highly skilled in cinematography and editing. When creating concept art, Roi often combines his pen and pencil skills with technical expertise to create an immersive transmedia experience for the audience. Roi is also part of Media Studies faculty at Vigo University in Spain.

Claire: age 49, 25 years in the industry.
Location: Los Angeles, United States.
Current position: Freelance Director.

Claire directs high budget comedies, and most of her work ends up streaming on Netflix and Amazon Prime. Furthermore, Claire S. is the only professional with experience in content production for now offline application Quibi that I could reach, which made her comments and insights highly relevant for the research.

Ernabel: age 55, 30 years in the industry.
Location: New York, United States.
Current position: Department Chair of Communication and Media Culture Department at Saint Peter's University, New Jersey; Host of CUNY-TV's magazine show 'Asian American Life'.
Ernabel has broad academic knowledge and decades of experience as on-camera talent. Being a representative of the older generation of TV producers, she always spends significant amount of time in research, and once the segment is filmed, she closely collaborates with editors. Most of Ernabel's television experience pre-dates the contemporary standard of multitasking content creators.

The interview sessions lasted between 30 to 50 minutes. During the conversations, I relied on the prepared questions, starting with a set of inquiries about the participant's educational background and professional experience. The second set probed techniques that proficient crew used while filming vertical videos and principal aesthetic differences between widescreen and vertical content. The interviewees also deliberated changes in their production workflows, conditioned by the necessity to produce specialised or re-purpose the existing footage for multiple distribution outlets. The last set of questions investigated the benefits and limitations of a smartphone as a consumer recording device and its integration into a standardised professional workflow. The data was transcribed with the help of AI powered speech to text application Otter. I examined transcripts using thematic analysis, inductive focus coding and manual memos, as a good fit for the investigation of a limited amount of non-numerical data such as six interviews. According to Virginia Braun and Victoria Clarke, 'An inductive approach to data coding and analysis is a bottom-up approach and is driven by what is in the data. This means the codes and themes derive from the content of the data themselves—so that what is mapped by the researcher during analysis closely matches the content of the data' (Braun and Clarke, 2006, p.58). Furthermore, for a researcher with limited experience in qualitative data sets, the thematic analysis provided enough flexibility as I was able to organise data into broader themes before focusing on particular comments and observations.

I was also aware of the limitations of thematic analysis, understanding that I bring my own set of values into the data as a researcher—starting with the choice of participants and ending with the context where the data was presented. Lastly, I consider potential limitations coming from a smaller data set, but in exchange, I relied on the fact that the diversity in skills, professional focus, and geographical location of interview participants will improve the relevance of my sample. As Lawrence Neuman stated in his book *Social Research Methods Qualitative and Quantitative Approaches* (2013), in qualitative research, 'it is their relevance to the research topic rather than their representativeness which determines the way in which the people to be studied are selected' (Neuman, 2011, p.247).

Chapter Three: Methodology – Creating Artefacts

The last practical methodological step was the production of original single-take widescreen and vertical music videos featuring a solo artist and a group of performers, but due to the disruptions caused by the COVID 19 pandemic, the time allocated for the interpretation of interview data overlapped with the filming schedule. By the time I needed to start with the production of thesis artefacts, I had finished a preliminary analysis of transcripts and worked out some of the key themes that came out of the interviews, but because of the delay, I couldn't test all applicable observations from interview participants in the filming component. Nevertheless, I completed the thematic analysis in parallel with the filming, which allowed me to use interview insights during post-production and critically reflect on my workflow.

The subject of the single performer music video was Mr. Russel Henderson, artefacts #1 and #2. He studied at Leeds College of Music and has had a twenty-year-long career as a performer in the United Kingdom and New Zealand. In search of data assortment with a single instrument, Mr. Henderson agreed to play two different original scores with alto and baritone saxophone. Looking to film multiple subjects in artefacts #3 and #4, I approached local Wellington ensemble Ras Judah & the Culture Embassy. Culture Embassy is a collective of musicians featuring multiple instrumental groups and vocal performers. The band's line-up is never firm because it depends on the members' availability and the list of songs musicians collectively agreed on before each gig. Filming a ten-piece Pan African band during hi tempo music numbers where accents quickly switch between the instruments cannot be more different from the stage atmosphere created by a single saxophone player. Such disparity between the subjects gave me an excellent opportunity to explore expressive possibilities of widescreen against tall aspect ratio while filming two radically different performances. Furthermore, by alternating smartphone orientation from vertical to widescreen after each take, I was in the position to observe and document the benefits and limitations of both under the similar subject. Lastly, I organised solo and group filming sessions in the studio environment to allow myself a greater degree of creative control and minimise possible external disruptions expected when filming in venues with a live audience. I will further discuss findings from this methodological step in chapter six.

To comply with the requirements outlined in my approved ethics application, I informed all participants and the studio crew that the primary reason for my filming was creating research data. However, both parties strongly expressed interest in using the final videos, particularly vertical versions, on their social media channels. I agreed to share the videos once my research was finished. To budget studio time but still get the variety in data, I arranged with the performers to film four different tracks with up to four takes per title. As expected, the first take of every number was plagued with technical and coordination problems, but by the time we started the second or third slate musicians were synced, and I had a solid understanding of the performance timeline. Whenever possible, I read visual cues from the musicians and led the camera in advance where the action was. Between two studio sessions and eight music numbers, I filmed a total of 32 takes, lasting between three to seven minutes. For the study, I disregarded all outtakes, technical interruptions, and false starts and ended up with 20 complete takes for analysis and reference.

For the fifth and sixth artefact, I needed original widescreen material filmed with a professional camera, with enough variety to create a cinematic narrative and edit a vertical trailer from it. A kind of material made during independent film production aimed for dissemination through Netflix. As a source of the footage, I used the material from one of my recent freelance projects, a short film *Loss* (2019). To produce designated thesis items and explore footage re-purposing processes, I firstly edited a ten minutes and 33 seconds long widescreen version from approximately two hours of raw files. Once the director David Lee approved the cut, I re-edited the second version of the film in 1:2 aspect ratio that became thesis item #5. Lastly, I selected a range of moments from artefact #5 suitable for editing of artefact #6, a 30 seconds trailer resembling the style of Netflix vertical previews. Here is the summary of items I submitted with the exegesis. I also included a hyperlink to the unlisted media at YouTube, where you can watch the files, as I stay compliant with the platform copyright rules.

Item #1 Saxes in 722 V - A single take vertical music video featuring a solo talent.
Item #2 Saxes in 722 W - A single take widescreen music video featuring a solo talent.
Item #3 Mumandi V - A vertical single take music video of a multi-piece band.
Item #4 Mumandi W - A widescreen single take music video of a multi-piece band.
Item #5 Loss Vertical Film - A short vertical fiction film, edited in 1:2 aspect ratio from the original 16:9 widescreen material.
Item #6 Loss Vertical Trailer - A 30 seconds vertical trailer, in 1:2 aspect ratio.

The selection of *Loss* (2019) as a source of the original footage for creating thesis items #5 and #6 proved to be an excellent choice, since I could discuss creative and technical decisions from the beginning and end of the production workflow. From the role of cinematographer, I reflected on the benefits and limitations of a large camera body as a principal recording device against a smartphone I

used to film artefacts #1, #2, #3 and #4. At the same time, from the position of editor, I could highlight shot-by-shot differences between the aesthetics of the original widescreen and vertical version with great precision and elaborate on connotative meanings brought into the similar mise-en scene by different aspect ratios. More particulars on the filming of the thesis items #1, #2, #3, and #4 are available in chapter six, while I unpack the discussion on the re-purposing process and introduce items #5 and #6 in chapter seven.

Chapter Four: Tall Music Videos & 1:2 Film Trailers – Introduction

In this chapter, I dissect the dominant two and three-dimensional image parameters responsible for the building of composition in key texts sampled from YouTube VEVO and Netflix, as well as methods creators of video and cinematic content use to direct the audience's gaze. Furthermore, I compare how the interplay of similar mise-en-scene elements renders a different meaning in the widescreen and vertical versions of music videos and how cropping of widescreen footage affects the perception of Netflix vertical previews. Acknowledging cinematic language has evolved with the advent of technology, I also explore how footage selected for the construction of a condensed editing sequence often found in music videos and film trailers has to meet a few additional prerequisites if the distribution of a final product is aimed at a mobile screen. Some of the more apparent changes in that evolution were amendments in the screen aspect ratio, but, despite a significant difference between the width of the original 35mm 4:3 frame's and today's theatrical standard of 2.35:1, the language of moving images mainly stayed the same. In contrast, with proliferation of vertical content where the horizontal axis became shorter than the vertical one, the conventions of the language of moving images exhibited more profound changes. A summarise breakdown of dominant 2D and 3D characteristics of the sampled media texts gave me a set of data that backed my initial observations, that the vertical framing affected not only the aesthetics of the original footage featuring limited shot scale and more subtle camera movements, but also heavily influenced the editing style. Finally, as I dissect the key samples, I will challenge the idea of 'unedited footage' and highlight techniques the skilled crew use to preserved the perception of 'user-generated' content to an audience unfamiliar with professional production workflow.

In an effort to make an engaging music video aimed at a mobile audience, A-list artists combined the two mainstream filming styles commonly read by the audience as the most honest depiction of an event. That is, a single uninterrupted camera take, housed in a vertical frame. Such a creative approach is visible in five key vertical samples from YouTube VEVO #1, #2, #4, #7, #10, while others, if not filmed as single takes, exhibited significantly simplified mise-en-scene when compared with their original widescreen versions. Simplified mise-en-scene was commonly combined with fewer edit points, resulting in sequences with less separate shots. Furthermore, I noticed a camera movement unique to single take vertical music videos that I have never seen before during twenty years of my professional career: a repetitive dolly-in progress, followed by dolly-out pull initiated without apparent prompt in blocking. These camera movements were never aggressive in terms of pace, as they did not exhibit rapid changes of positions. The motion was fluid, with subtle starts and stops as the camera travelled at eye level, suggesting another person's point of view. The described movement is evident as principal camera motion in samples #01, #02 and #07. Here is the list of the key samples from YouTube and Netflix I analysed in this chapter. (YouTube sample - YTS, Netflix sample - NS). Similarly to the previous chapter, titles contain hyperlinks to media for the readers' convenience.

YouTube Sample #1 Mark Ronson - Nothing Breaks Like a Heart ft. Miley Cyrus / Vertical Version
YouTube Sample #2 Bebe Rexha - Last Hurrah / Vertical Version
YouTube Sample #3 Billie Eilish - You Should See Me in a Crown / Vertical Version
YouTube Sample #4 Taylor Swift - Delicate / Vertical Version
YouTube Sample #5 Billie Eilish – I Don't Wanna Be You Anymore, no Widescreen Version
YouTube Sample #6 Camila Cabello - Havana ft. Young Thug / Vertical Version
YouTube Sample #7 Katy Perry - Harleys In Hawaii / Vertical Version
YouTube Sample #8 Nicki Minaj - Chun-Li / Vertical Version
YouTube Sample #9 Sam Smith - How Do You Sleep? / Vertical Version
YouTube Sample #10 Sam Smith - Too Good at Goodbyes / Vertical Version
YouTube Sample #11 Halsey - Without Me / Vertical Version

Netflix Sample #1 Formula 1 - Drive to survive S1 / Vertical Version
Netflix Sample #2 Glow Up: Britain's Next Make-Up Star S1 / Vertical Version
Netflix Sample #3 Inside Bill's Brain: Decoding Bill Gates / Vertical Version
Netflix Sample #4 Marriage Story (2019) / Vertical Version
Netflix Sample #5 Mowgli: Legend of the Jungle (2018) / Vertical Version
Netflix Sample #6 Orange Is the New Black S1 / Vertical Version
Netflix Sample #7 Patriot Act (2018) / Vertical Version
Netflix Sample #8 Stranger Things S1 / Vertical Version
Netflix Sample #9 Sugar Rush S1 / Vertical Version
Netflix Sample #10 The Ballad of Buster Scruggs (2018) / Vertical Version
Netflix Sample #11 White Fang (2018) / Vertical Version

Besides the parallel that both rely on mobile viewers as a primary audience, another reason why I analysed vertical VEVO music videos next to Netflix vertical trailers was the fact that both formats tend to rely on condensed editing sequences to tell the story in relatively short running time, emphasise titles of key creative personnel such as directors, and use the audio track as the principal element of diegetic unity. Equally important, vertical music videos and vertical trailers serve a similar purpose of drawing the audience's attention to the principal widescreen piece. Through a detailed breakdown of listed examples, I will juxtapose how similar mise-en-scene elements are perceived differently in vertical and widescreen aspect ratios and why a tall frame commands a different approach to visual storytelling than the traditional widescreen. The Excel spread sheet containing a detailed breakdown of analysed parameters is available in appendix #4.

Chapter Four: Numbering the Shots

To reveal the trend in the shot scale and before commencing to quantitative analysis, I sorted and numbered every shot from the key samples, figures 36 (YouTube VEVO) and 37 (Netflix). The breakdown revealed that widescreen music videos had only 2.59 percent more medium or medium close-up shots and an 8.76 percent higher count of wide or full body shots than vertical. In contrast, vertical music videos comprise 11.61 percent more close-up and extreme close-ups than found in widescreen counterparts. The breakdown also revealed a complete absence of vista shots in both versions of music videos. Another important observation the summary in figure 36 revealed was a significantly higher number of edit points in widescreen than in the vertical videos. A fewer edit points in a sequence does not necessarily mean the scene features a simple camera movement or composition, but in the case of analysed music videos, a lower number of shots in the sequence was always coupled with a simple mise-en-scene.

YT music videos Footage breakdown	Total number of edit points	M/MCU	CU/ECU	W/F	V	Transitions shots and text
Widescreen	**1174**	770	148	232	0	24
		65.59%	12.61%	**19.76%**	0	2%
Vertical	**512**	322	124	57	0	9
		63%	**24.22%**	11%	0	1.8%

Figure 36: The distribution of shot scale in widescreen and vertical versions of key samples from YouTube VEVO.

The following figure 37 illustrates the spread of 2D and 3D traits I found in the sample. Weak magnetism was prevalent in both widescreen and vertical music videos, indicating a central position of the subject in the majority of shots. Another trait shared between both versions was the presence of strong directional vectors induced by the movement of the camera or subjects. The rest of the analysed qualities, figure-ground relationship, Z-axis articulation, presence of secondary framing, and direct address method had an even spread between the categories weak, detectable and strong, suggesting a great variety of compositions, camera movements, and styles of lyric deliveries in all samples.

YT music videos Summary of dominant 2D and 3D features		A: magnetism & asymmetry of the frame	B: dominant directional vectors	C: objects figure/ground relationship	D: Z-axis articulation	E: secondary framing	F: direct address method
Nothing Breaks Like a	W	Weak	Strong	Strong	Detectable	Strong	Strong
Last Hurrah	W	Weak	Strong	Strong	Detectable	Strong	Strong
You Should See	W	Weak	Strong	Weak	Weak	Weak	Weak
Delicate	W	Weak	Strong	Strong	Strong	Strong	Weak
Camila C. - Havana	W	Detectable	Strong	Weak	Strong	Strong	Weak
Harleys In Hawaii	W	Detectable	Strong	Detectable	Detectable	Weak	Weak
Nicki M. - Chun-Li	W	Weak	Strong	Weak	Strong	Strong	Strong
How Do You Sleep?	W	Weak	Strong	Strong	Strong	Detectable	Strong
Too Good At Goodbyes	W	Weak	Strong	Strong	Strong	Strong	Detectable
Halsey - Without Me	W	Detectable	Strong	Strong	Detectable	Detectable	Weak
Nothing Breaks Like a	V	Weak	Detectable	Strong	Strong	Strong	Strong
Last Hurrah	V	Weak	Strong	Strong	Strong	Strong	Strong
You Should See	V	Detectable	Detectable	Strong	Weak	Weak	Detectable
Delicate	V	Detectable	Strong	Strong	Detectable	Weak	Strong
I Don't Wanna be You	V	Strong	Strong	Weak	Detectable	Strong	Weak
Camila C. - Havana	V	Detectable	Strong	Detectable	Detectable	Strong	Detectable
Harleys In Hawaii	V	Weak	Strong	Detectable	Detectable	Detectable	Detectable
Nicki M. - Chun-Li	V	Detectable	Strong	Weak	Strong	Weak	Strong
How Do You Sleep?	V	Weak	Strong	Strong	Strong	Weak	Strong
Too Good At Goodbyes	V	Weak	Strong	Strong	Detectable	Weak	Strong
Halsey - Without Me	V	Weak	Strong	Strong	Detectable	Weak	Strong

Figure 37: Principal 2D and 3D aesthetic characteristics of sampled music videos.

Figure 38 describes the shot breakdown for widescreen and vertical Netflix trailers. Analysed trailers exhibited overall shot scale distribution comparable to music videos when it comes to medium, close-up, and wide compositions. Both versions of trailers had an almost similar number of medium and medium-close up shots, followed by ten percent more close-ups and extreme close-ups in a vertical and a slightly higher percentage of wide composition in widescreen versions. When it comes to the type of shots used to build the diegetic space, the trailers featured vista shots and significantly more transitions, such as fade to black and titles, than music videos.

Netflix trailers Footage breakdown	Total number of edit points	M/MCU	CU/ECU	W/F	V	Transitions shots and text
Widescreen	**988**	553	148	189	15	85
		55.97%	14.98%	**19.13%**	1.5%	8.6%
Vertical	**352**	197	89	48	2	16
		56%	**25.28%**	14%	0.5%	4.5%

Figure 38: The distribution of shot scale in widescreen and vertical versions of key samples of Netflix trailers.

Figure 39 summarizes 2D and 3D characteristic and reveals a dominance of detectable magnetism and asymmetry of the frame, suggesting that most compositions conformed to the rule of thirds and exhibited a clear frame intention in the composition. I.e. A principal subject or object of interest was not in the centre of the frame. Furthermore, frequent strong figure-ground relationships in the footage indicate a lack of tilted camera angles, while the absence of moments where the protagonist addresses the lens directly shows the author's desire not to break the fourth wall or rupture created diegetic space. The only exception in filming and editing style was a trailer for the stand-up comedy show *Patriotic Act*, which exclusively used the direct address method for its entire running time. The rest of the characteristics describing the presence or absence of directional vectors, Z-axis articulation, and secondary framing had a full spread from weak to strong, suggesting the filming style was determined by topic and subject, same as in the case of music videos.

Netflix trailers Summary of dominant 2D and 3D features		A: magnetism & asymmetry of the frame	B: dominant directional vectors	C: objects figure/ground relationship	D: Z-axis articulation	E: secondary framing	F: direct address method
F1 - Drive to Survive	W	Detectable	Strong	Strong	Detectable	Detectable	Weak
Glow Up	W	Detectable	Detectable	Strong	Weak	Detectable	Detectable
Inside Bill's Brain	W	Detectable	Detectable	Strong	Detectable	Detectable	Weak
Marriage Story	W	Detectable	Detectable	Strong	Weak	Detectable	Weak
Mowgli	W	Detectable	Strong	Detectable	Strong	Detectable	Weak

Orange is the New Black	W	Detectable	Detectable	Strong	Detectable	Detectable	Weak
Patriot Act	W	Detectable	Detectable	Strong	Detectable	Weak	Weak
Stranger Things	W	Detectable	Strong	Strong	Strong	Strong	Weak
Sugar Rush	W	Detectable	Detectable	Strong	Detectable	Weak	Weak
The Ballad of	W	Detectable	Detectable	Strong	Strong	Detectable	Weak
White Fang	W	Detectable	Strong	Strong	Detectable	Detectable	Weak
F1 - Drive to Survive	V	Detectable	Strong	Strong	Detectable	Weak	Weak
Glow Up	V	Weak	Detectable	Strong	Weak	Weak	Detectable
Inside Bill's Brain	V	Detectable	Detectable	Strong	Detectable	Detectable	Weak
Marriage Story	V	Detectable	Detectable	Strong	Detectable	Detectable	Weak
Mowgli	V	Detectable	Strong	Strong	Detectable	Weak	Weak
Orange is the New Black	V	Detectable	Detectable	Strong	Weak	Weak	Weak
Patriot Act	V	Weak	Weak	Strong	Weak	Weak	Strong
Stranger Things	V	Detectable	Detectable	Strong	Detectable	Detectable	Weak
Sugar Rush	V	Detectable	Strong	Strong	Detectable	Weak	Weak
The Ballad of	V	Detectable	Strong	Strong	Strong	Detectable	Weak
White Fang	V	Detectable	Strong	Strong	Weak	Weak	Weak

Figure 39: Principal 2D and 3D aesthetic characteristics of sampled trailers.

Lastly, the number of edit points wasn't directly comparable in the case of trailers because of significantly different running times. The widescreen version of the trailers runs for 107 seconds on average, compared to 30 seconds for vertical ones. Yet, if the combined running time was divided by the total number of shots in each group, I concluded there was no significant difference in the editing pace between widescreen and vertical trailers. In contrast, music videos had almost identical running lengths but significant differences in the number of edit points, 1174 in widescreen and 512 in vertical. I would argue a 57 percent difference in the number of shots used to illustrate the same length of music videos indicated another aesthetic approach to storytelling in a vertical frame from one observed on widescreen. In appendix #5, I have included an alternative arrangement of figures 37 and 39, with titles of key widescreen samples next to their vertical counterparts and colour-coded highlights of trends observed in the footage.

Chapter Four: Tall Music Videos and Their Widescreen Counterparts

The music video *Mark Ronson - Nothing Breaks Like a Heart ft. Miley Cyrus* is included first in the analysis as a prime example of differences between official widescreen and vertical music video of the same title. Released as a single in late 2018, a widescreen version of this music video features a dozen scenes intercut in parallel with the primary car chase action. Such editing approach produced a dense visual sequence designed to highlight multiple pressing social issues in American society. The filming involved closing the 1.1km six-lane New Darnytskyi Bridge (Kyiv/Ukraine), a helicopter and dozens of other vehicles, stunts, multiple camera crews and large groups of stand-ins. The video also contains multiple self-referencing elements and homages to early works of the singer that are irrelevant for this study, but overall contributed to intricacy of mise-en scene.

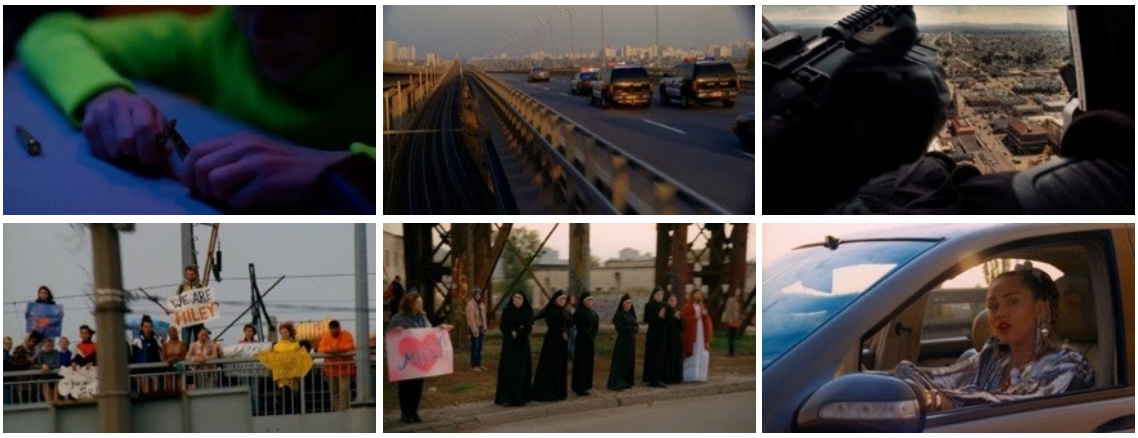

Figure 40: Mark Ronson - Nothing Breaks Like a Heart ft. Miley Cyrus. Selected screen shots from the range of scenes.

An opening shot introduces a young girl playing with two high calibre bullets in each hand, her interaction with the bullets can be read as playtime akin to two dolls talking to each other. From this comment on accessibility of weapons in the US, Cyrus and Ronson move on to critique of ‘stan’ culture with multiple scenes of people holding signs with the singer’s name, (figure 40). A USA Today article from September 2021 highlighted the growing problem and the dangers of stan culture. Members of stan groups are commonly portrayed as fans, who are willing to go to any length to prove their devotion to the celebrities of their choosing, but their behaviour can lead to addictive tendencies and stalking behaviour (Oliver, 2021). Stan scenes are followed by helicopter shots of slow police highway pursuit after a single vehicle and details of automatic weapons. Such images can be linked to controversial 'VIP pursuits' as they resemble the iconic OJ Simpson chase from June 17th, 1994, reminding the audience of controversies associated with the lenient interpretation of law when it comes to celebrities, (figure 40).

Figure 41: Mark Ronson - Nothing Breaks Like a Heart ft. Miley Cyrus. Selected screen shots from the range of scenes.

The scene starting at the 25-second mark shifts focus to Catholic church scandals as the audience can see a line of nuns standing by the side of the road together with 'Jesus', but the traditional representation of religion is then juxtaposed with a sequence of clergy in a strip bar, that culminates with a CU of a priest gazing at half-naked dancers, (figure 41). Scenes highlighting several other pressing social issues are addressed during the second half of the video, as the protagonist's car passes by a shooting range, a pair of girls posing for a selfie, a gay couple, and a row of kneeling American NFL players before it becomes stuck amid a riot caused by a 'Black Friday' purchasing spree. The music video ends with a montage of the violent car crash, fans rushing to meet the celebrities, and a crucifixion effigy of the main protagonist in front of a red carpet, (figure 41). In summary, the widescreen version highlights multiple pressing issues of first-world countries, with a strong reference to the United States. Still, none of that political charge is visible in the second official release of *Nothing Breaks Like a Heart.*

Figure 42: Mark Ronson - Vertical version of Nothing Breaks Like a Heart featured only one scene and a single corrective camera pan.

In contrast to the complexity of the first version, the vertical video of the same song contains a single scene filmed as one uninterrupted take, (figure 42). A strip club environment is a principal mise-en-scene component that together with the singer's wardrobe creates a sense of spatial unity between the widescreen and the vertical version. Fifteen dancers walking in circle on stage present a backdrop for the singer while repetitive choreography produces a sensation of a conveyor belt, reminding the audience of the problem of exploitation and over sexualization of women. During the song bridge at two minutes 34 seconds, the main protagonist singer Mylie Cyrus breaks the gaze with the lens as a subtle six seconds long corrective pan opens the space for Mark Ronson. When the camera settles, the singer re-establishes the eyeline with a direct gaze at the camera, whilst two protagonists share the central position in a balanced frame until the end of the video.

The lyrics of the song stayed the same in both versions yet the images used to illustrate them were so dissimilar, which prompts the question of why the creative approach was so different between the aspect ratios? I would argue the widescreen version was created as a provocative statement targeting emotions of a broad audience by covering topics relevant to multiple ethnic and marginalised social groups. Moreover, most of the scenes in the widescreen version contain a complex multi-layered composition, that combined with primary and secondary motion would challenge the perception of mobile audiences and lose their intricate details on a small screen. As an alternative, a simplified mise-en-scene of the vertical version filmed at a single location produces more hermetic diegetic space leaving fewer options for lyrics interpretation, but at the same time features almost uninterrupted eye contact between the singer and the audience. I would argue the emphasis on the aesthetics of the direct address method is instrumental in capturing the viewer's attention without tertiary motion and vital to re-enforce the feeling that the artist is talking directly to the audience. Lastly, despite the fact it is envisioned as a single take, the video contains one cut at the 85th second, hidden behind the lens flare. The reason for the edit point might be the talent's desire to combine performance quality of two different takes into one, but the fact the cut was masked with a subtle visual effect suggested the desire of the director Pierre Dupaquier to preserve perception of an uninterrupted take in the audience's mind.

Bebe Rexha's Last Hurrah (2019) was published two months after *Nothing Breaks Like a Heart*, and it follows the same creative pattern of two fundamentally different official releases. The widescreen version of *Last Hurrah* contains multiple scenes intercut together, as the story once again portrays the issues of violence against woman, sexual scandals of the Catholic church and aggressive behaviour of social media trolls, (figure 43). Another recognizable parallel with the *Nothing Breaks Like a Heart* is the use of backdrop architecture and wardrobe, in support of the notion of spatial unity between the two

Figure 43: A selection of screenshots from Last Hurrah.

versions. The widescreen version also features more explicit footage as well as product placement[x] of headphones and cars (Beat and Alpha Romeo).

Figure 44: Starting and ending (dolly +/-) camera positions from the vertical version of Last Hurrah.

In contrast, the vertical version again follows the trend of a single take, featuring much simpler blocking and lower production value, (figure 44). Equally important, as a perpetuation of already established conventions associated with vertical video, the lack of edit points suggested an unaltered, thus the more honest account of the event. An appearance of a single centrally positioned protagonist in an angel costume, singing alone in the narrow side street, produces a more subtle interpretation of the lyrics than the original. Where the widescreen version relies on montage to create connotative meanings, the *Last Hurah's* vertical version relies on camera movement and variation in performance intensity to establish the arc of the story. The video starts with a ten seconds long dolly during the first verse, as the camera moves away from an M shot to a W. Once the camera comes to a stop, it almost immediately reverses the movement to resume the original position as the first lyrics are heard. This sequence is repeated five times (five dolly-in and five dolly-out) until the song reaches the bridge in the melody. Because the singer delivered the last verse and crescendo with more intensity, the camera travelled at double speed, covering the distance from W to M shot in five seconds and making sure the last dolly plus action finishes on the final note of the song.

While gradually changing the point of view, the forward motion of the camera increases the size of the subject and its overall perceived importance in composition. Furthermore, the proximity of the subject allows the audience to read more details on the object of interest and emotion on a face as one of the predominant factors of attention synchrony. According to David Bordwell and Kristin Thompson:

> Our eyes and brains are better suited for noticing differences than for concentrating on uniform, prolonged stimuli. Thus aspects of mise-en-scene will attract our attention by means of changes in light, shape, movement, and other aspects of the image (Bordwell & Thompson, 2001, p.158).

That is, if the camera pulls away from the subject such movement signals de-escalation of tension or watering down of the emotion, unless the new position reveals another story-driving element. With the exception of the second movement synced with the start of the lyrics and last verse that singer delivers with more intensity, there are no other strong accents in blocking or unique variations of choreography in *Last Hurrah*. Furthermore, the camera does not adjust the height or angle, neither does the change of position reveal new story driving elements. The decision to use the same secondary motion in such a repetitive manner is questionable, because 'unmotivated camera moves are often labelled as unmistakable stamps of amateur video or film due to the fact unmotivated secondary motion draws attention to itself, instead of the event that footage is supposed to cover' (Zettl, 2016, p.295). So, where does the motivation for camera movement come from?

I suggest that a completely static vertical version of *Last Hurrah* would suffer from the lack of refreshing visual features able to capture audience attention for the entire duration of the track. For that reason, the director Joseph Kahn, decided to introduce the additional camera movements as compulsion to keep the mobile audience glued to the screen with the strongest of directional vectors, a motion vector. As Smith noted:

> Sudden onset of motion, such as forward movement of the camera or appearance of faces leads to attentional synchrony. Smith also concluded: that camera dollies illustrate how viewers must reconcile the bottom-up audio-visual factors, fighting over their attention with their own top-down interest in order to decide where to attend (Smith, 2011, p.19).

That is, when the camera moves directly toward the centre of the frame, an already dominant singer's face gets even more attention in an M shot as details of facial expressions are gradually revealed. Once reversed, a W shot encompasses the entire body together with prop wings and a slit in the skirt, creating

multiple points of interest for the audience. I would conclude both music videos feature multiple camera movements, but the vertical version of this pair relies much more on motion vectors as the primary driver of attention synchrony due to its minimalistic composition.

Unlike *Nothing Breaks Like a Heart* and *Last Hurrah*, other than the presence of a single motorcycle, there is no similar wardrobe or scenery to convey the sense of unity between the two versions of *Harleys in Hawaii* (2019), (figure 45).

Figure 45: Widescreen version of Harley in Hawaii features multiple scenes and rich variety of camera angles.

Figure 46: The vertical version of Harleys in Hawaii contains a variation of repetitive dolly +/- camera movement described in Last Hurrah.

Instead, this vertical music video relies on a variation of camera motion described in *Last Hurrah* to improve the attention synchrony within a single take, (figure 46). The *Harleys in Hawaii* vertical video starts with a 34-second long static MCU shot, before the camera commences the 32-second dolly out movement on the first chorus, changing the composition to a W shot. After the initial movement, the camera stays in place for eight seconds before doing another very slow 50-second long dolly-in movement back to the original MCU. As the camera pulls back to W shot, the singer's blocking is off centre, producing the frame intention towards the right but still yielding a very stable composition. During the song bridge and right before the last chorus, the camera tracks back again over the course of 43 seconds and lands at a W shot seen before. Like in the previous case, I would argue that the principal reason for repetitive dolly camera movements comes from a lack of other mise-en-scene elements able to produce tension in the shot. As Zettl noted, 'the z-axis index vector, such as a person looking or pointing directly at the lens, has a very low magnitude in comparison to an object that is moving toward the camera' (Zettl, 2016, p.123). In conclusion, despite the fact that the camera movements were not strongly motivated by script or blocking, the changes of framing and the dominance of motion vector produce a sensation for the spectator that is more engaging than a static composition.

The next title, *Billie Eilish I Don't Want to be You Anymore* (2018) is the only key example released as a vertical video only, but that is not the only reason why it was important to dissect this music video. Eilish published the song originally on Spotify in December 2017, but the company also required a music video from the artist. A few weeks later in January 2018, the track became available on YouTube VEVO, but no official widescreen version was ever issued. Without large commercial success up to that point, Eilish was still an emerging artist with limited budget for a large-scale production. For that reason, she opted for a vertical music video targeting mobile audiences and minimalistic design that saved on production costs. Testimonials of Alyssa Bereznak from 'The Ringer' blog provides invaluable insights into behind the scenes circumstances during the filming of *I Don't Want to be You Anymore*. Being one of the pioneers of vertical music videos, the director Eli Born was hired for the project. In the past Born was filming vertical material as additional footage while recording principal widescreen photography, but this was the first-time vertical video was the primary event. After the shoot, the director said 'the production of the vertical video was both effective and budget-friendly

because it required only presence of Eilish, himself, one lighting person, and a camera operator' (Bereznak, 2019). Furthermore, the director described vertical framing as beneficial for his aesthetic approach, as he could use the lens flares produced by close proximity of the light sources, without risking revealing them in the composition, (figure 47).

Figure 47: A selection of screenshots from Billie Eilish I Don't Want to be You Anymore (2018)

Because it was released only as a vertical video, *I Don't Want to be You Anymore* is aimed primarily at mobile audiences, but that had some trade-offs for watching the video on widescreen. Having a very minimalistic production design, visual dynamics in this music video were created through changes in camera angles, variety in shot scale and lighting conditions. If this vertical video is played on a widescreen, the viewing surface will incorporate pillar boxes matching the dominant colour of the production design, (figure 48). This was a fresh solution in contrast to traditional black pillar boxing, as similar colouring challenged the dominance of principal footage to a lesser extent. Lastly, well-defined vertical composition, with space available only for the protagonist and her reflection, created another layer of subtext that works well for the story of a person struggling with the depression Eilish tells through the song lyrics.

Figure 48: A selection of screenshots from Billie Eilish I Don't Want to be You Anymore (2018), with matching colour of pillar boxing in widescreen.

The next sample, Taylor Swift's *Delicate* (2018), further demonstrates how professionally created music videos use encoded perception of the vertical frame to engage the audience. In contrast to the elaborate mise-en-scene and rich choreography of the official widescreen version, (figure 49), a single continuous vertical take filmed on a seesaw in an ordinary park, brings to mind a light-hearted atmosphere of daydreaming that many users can connect with, (figure 50).

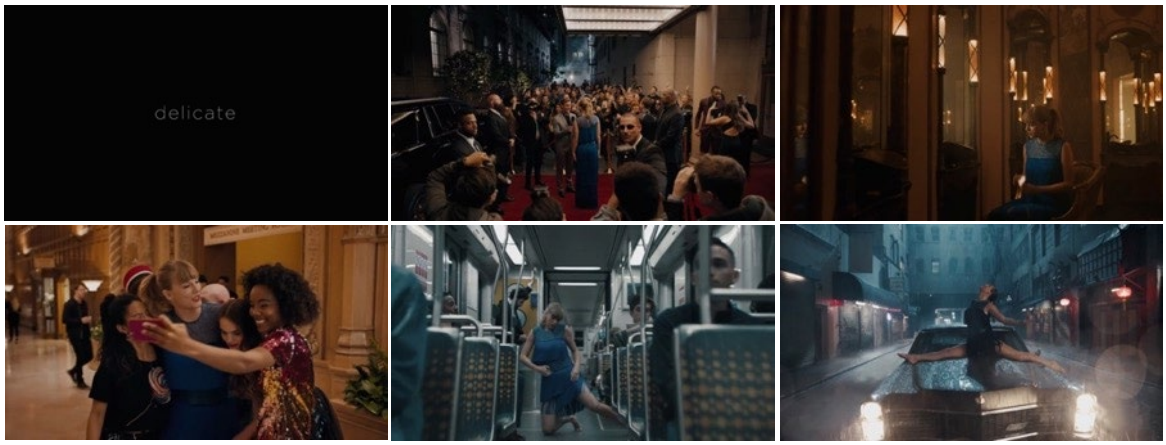

Figure 49: Selection of screenshots from widescreen version of Delicate (2018).

Figure 50: Selection of screenshots from vertical version of Delicate (2018).

Even though the footage mimics user-generated smartphone video aesthetics, there are multiple clues that signal the material was produced with help of the crew, rather than the artist alone. First, during all vertical and horizontal movements, some of them sudden and fast, the camera stayed fixed to the central image axis suggesting strong rigging of the camera support or presence of an advanced Steadicam system. This observation is further confirmed with several moments in which the protagonist has both hands in the air, thus not being able to hold her phone or another recording device. Second, the high quality of audio, with clean voice and no interference from any exterior elements, signals a dubbed track recorded in a studio. If Taylor Swift as a globally established artist looks to hide post-production elements in her music video, the next logical question is why? I would conclude the framing style of a selfie in combination with controlled amount of footage imperfections, such as inconsistent lighting and the use of natural sources such as the Sun, conveys the notion of user-generated content, while the audience consumes a highly post-produced single take music video. The use of a direct address method further fuels the illusion due to a breaching of the fourth wall as the singer wishes to establish a closer emotional connection with the audience. According to Zettl:

> the performer's demeanour, rather than a specific production technique, plays a big part in establishing ground for a variety of direct-address messages. Film is quite different in this respect as film actors do not commonly break the fourth wall, and because of that the audience does not see them as a part of daily life (Zettl, 2016, p.237).

That is, by doing a mundane action such as a seesaw ride, Taylor Swift wants to be viewed as an ordinary young woman in contrast to the connotation of a celebrity persona who wishes to hide from the world of fans and paparazzi depicted in the widescreen version of *Delicate* (2018). It is safe to say, the audience unfamiliar with a workflow of a professional music video set wouldn't be able to identify the camera was stabilised or recognise subtle colour grading[x] of the footage, but I would argue *Delicate* (2018) relies on vertical framing more than any other aesthetic or technical aspect to persuade the audience into the authenticity of the depicted moment.

Figure 51: Selection of screenshots from vertical version of You Should See Me in a Crown (2018).

Following the positive reception of her previous minimalistic and affordable music video, *I Don't Want to be You Anymore*, Eilish released another single, *You Should See Me in a Crown* (2018), in mid-2018 as a vertical video only. In this track, Eilish conveyed a story of metamorphosis from human to spider on a single colour infinite studio backdrop, with no changes in camera angle but through vivid variety of shot scale and camera heights, (figure 51). A slim, well-defined 9:16 composition, occupied mostly by the protagonist's body, once again contributed to the sensation of hermetic space and conveyed the notion the scene might be unfolding inside the protagonist's mind. A new music video for the title was published nine months later, and despite the fact it was released as a successor, it followed the tendency of more complex mise-en-scene commonly associated with widescreen versions, (figure 52).

Figure 52: Selection of screenshots from the widescreen version of You Should See Me in a Crown (2018), directed by critically acclaimed contemporary artist Takashi Murakami.

While looking for the data on viewing numbers and posting dates of the two versions, I noticed a previously hidden trend, that once put in context, became one of the key takeaways of this research step. In April 2019, when the widescreen version of *You Should See Me in a Crown* was released, the engagement potential of vertical video within a mobile audience was already apparent, but only when I listed posting dates for all key texts sampled from YouTube VEVO in the same Excel sheet, I noticed most of the titles analysed in this chapter followed the approximately three to four week delay between the first and second version of the music video. So what was the reason behind such posting trend?

Commonly, the first music video released for a new title was widescreen. As a second release, the vertical version often got only 1-10 percent of the widescreen's viewing numbers and the disparity was expected because the buzz surrounding the vertical video was always lower; after all, the video might be new, but the song was old. However, comparing the numbers for *Billie Eilish I Don't Wanna Be You Anymore* (265 million), and *Billie Eilish - You Should See Me in a Crown* (237 million), both initially released as a vertical video, I found the tracks has similarly high numbers as any widescreen original releases. Eilish and Murakami planned to make an official widescreen version for *You Should See Me in a Crown* early on, but due to the complexity of production it took Murakami nine months to finish elaborate animated video, and by that time, the song was already dated in the audience's mind. As a result, despite the growing popularity of Eilish and a much higher budget for the widescreen version, the video garnered only a third (85 million) of the views compared to the minimalistic vertical video. Such a significant difference between the two versions brings to the forefront a question of what primarily influences the popularity of the video: the choice of an aspect ratio and footage used to illustrate the lyrics, or which one is released first?

It would be speculative to discuss the popularity of one artist against the other, but the fact some videos gather unusually high numbers of YouTube hits can possibly be attributed to being pioneers of a genre. Despite the fact *Havana* (2017) was released as an alternative version of an already published widescreen video, 143 million views (as of June 2022) represents by far the largest number of any vertical video I encountered during the analysis. I couldn't determine with certainty what the reason was for that unusual success, but Camila Cabello was one of the first A-list artists to release an official vertical video back in 2017, and arguably was one the first to establish the three to four weeks posting trend between the versions, (figure 53).

Figure 53: Selection of screenshots from vertical version of Havana.

The occasionally tilted hand held camera simulates the framing of an untrained operator, or a shooting style of another subway car commuter, as the story follows three friends in a good mood traveling on the train. The equally successful widescreen take of this song is produced as a collaboration, *Camila Cabello - Havana ft. Young Thug* (2017), and is another representative of a traditional high budget music video production, but there is another reason why this text is highly relevant for the analysis in this chapter. Besides apparent differences in budget, the widescreen music video of *Havana* shows how changes of aspect ratio within a single sequence carry a strong connotation in itself, as our perception subconsciously encodes meaning to a particular style of the footage and shape of the frame, (figure 54).

Figure 54: Selection of screenshots from widescreen version of Camila Cabello - Havana ft. Young Thug (2017).

The widescreen version of *Havana* starts with a short scene where exaggerated blocking, over the top actors' reactions, and flat lighting resembles the ambience commonly associated with soap operas. After 40 seconds of running time, the audience experiences an aggressive cut as the image transforms from 16:9 wide aspect ratio into elongated letterboxed 2.35:1 frame. The appearance of the letter box mask signals the switch from video to cinematic space, and more elaborate lighting design supports that notion. A sense of spatial unity is preserved as the actress continues her action in the secondary frame of the new location, but the onscreen atmosphere is quite different, with less saturated colours and pools of light separating positive volumes. The screenshots on the left and in the middle, top row in figure 54, show the transition moment. Harper Cossar (2011) describes this editing method as rupture of screen

space, as the shift in aspect ratio cues the viewer that a different experience is now taking place. 'Viewers recognise the letterbox aesthetic as remediated film viewing, and are cued via the aspect ratio shift to consume rather than participate' (Cossar, 2011, p.239).

From the moment of the switch, the *Havana* music video has a 109-second intro before the first tunes of the song are heard, and after six minutes and 42 seconds, the video ends with a title card saying, 'This is Dedicated to the Dreamers', the last screen shot in figure 54. The widescreen release of *Havana* follows a trend of high budget music videos featuring actors, YouTube celebrities, sport stars and multi-layered meanings that raise awareness of some of the major global issues such as racial and social injustice. The closing sentence resonates with the entire immigrant population of the United States, but speaks directly to 800 thousand undocumented children, born to parents who came illegally into the US, colloquially named 'dreamers'. In contrast, the dynamics between performance, camera angles and cuts, together with multiple secondary frames visible in the vertical version of *Havana* creates an engaging visual sequence, but the vertical video doesn't have any resemblance to the elaborate widescreen production and it doesn't carry any strong political messages.

The next video I analysed, Sam Smith's *Too Good at Goodbyes* (2017), is another example of how the rupture of screen space, combined with a coded perception of the vertical frame, can be used to convey the feeling of authenticity in the audience mind. With this title, Sam Smith established a style of including a behind-the-scenes moment as an opener of his music videos, a trend recognizable in several of his later works. During the opening 30 seconds of widescreen version of *Too Good at Goodbyes* (2017), the singer is filmed with a face looking down as the voices of the crew getting ready can be heard in the background, (figure 55).

Figure 55: The opening sequence of the widescreen version of Too Good At Goodbyes (2017).

Figure 56: Too Good at Goodbyes - The opening composition and title card of vertical video.

Similarly the vertical version of *Too Good At Goodbyes* opens with Sam waiting while the audience hears the exchange between the assistant director and camera team, confirming everyone is ready for the take and video monitoring is available for other departments present on site. While he listens to the conversation Sam breaks the gaze with the lens and glances off-camera, presumably awaiting the go-ahead from the director, but instead of a final 'action' instruction, the title card breaks the wait, (figure 56). Such false starts are a common occurrence on any video or film set, and the cause of the delay could be anything from a crew member that is not settled in position to a passing siren picked up by microphone to name a few; what is interesting is the fact that these pre-roll moments made the final cut and once again, suggest a significantly different aesthetic approach to vertical and widescreen storytelling.

Figure 57: Selection of scenes from multiple locations featured in the widescreen version of Too Good at Goodbyes.

Both versions of the *Too Good at Goodbyes* music video are directed by the same person and feature an impromptu gesture, but the question arises as to why the director Luke Monaghan filmed 20+ actors at 16 locations in the widescreen version (figure 57) and opted to tell a story with a piece of pre-roll footage and single take in vertical one (figure 58)?
I wasn't able to uncover why the vertical version was filmed in black and white technique and widescreen in colour, but as in the case of *Delicate*, I would conclude a simpler mise-en-scene, and a single take video conveys the feeling of authenticity as an unedited account of the event stronger than widescreen version. To support that notion, the visual dynamics within the single vertical take of *Too Good at Goodbyes* were created through a combination of singer and camera movements, not with edit points. As the song develops, the camera floats between variation of M and CU shots, changing the distance slowly as the singer occasionally break the gaze and re-engage the camera with a direct address method. Unlike highly choreographed matching widescreen shots, with no colour and simple blocking, the vertical version produces a more subtle and intimate account of *Too Good at Goodbyes* lyrics. Lastly, out of all the texts I analysed within the scope of this research, this track exhibited the longest period of 18 months between the two official releases. In the observed trend on the YouTube VEVO channel, a longer waiting period means lower viewing numbers for the second release, and that is evident when 1.26 billion hits achieved by the widescreen version of *Too Good at Goodbyes* are compared to only 2.5 million views accumulated by vertical video. Such a stark difference is also a testament that in a saturated media market, even an artist at the level of Sam Smith, able to produce a global music sensation, has to follow the publishing pattern for his music videos if he wants to achieve a higher viewing number for the alternative release.

Figure 58: The two most dominant compositions found in a single-take vertical version of Too Good at Goodbyes, tight M and CU shot.

Continuing to harvest the potential of behind-the-scenes moments to demystify professional production, Sam Smith started another high budget video, *How Do You Sleep Tonight* (2019), with a scene showing the edge of 'infinite studio background', scores of cables, century stands, and video village monitor; a site commonly reserved for a production crew only. As the lyrics revolve around a person disappointed in an emotional relationship, the motionless body of the singer is juxtaposed with the busy studio environment, once again looking to equalise the feelings of an on-screen star performer with the emotions of a 'real person'. The rest of the video leads the audience through several elaborate scenes involving very complex props and decor intercut with a principal choreography, (figure 59).

The vertical version plays on similar emotion as it opens with a slow dolly shot of an inert protagonist, now sitting in front of the dark infinite backdrop, but without any extravagant scenery present in original widescreen, the lyrics were illustrated in a different manner, (figure 60).

Figure 59: Selection of screenshots from widescreen version of How Do You Sleep Tonight (2019).

Figure 60: Selection of screenshots from vertical version of How Do You Sleep Tonight (2019).

Relying only on a single camera angle, the director explores the potential of Z-axis articulation and the longer vertical edge of the frame to create visual dynamics. For that reason, the mise-en-scene is dominated by matching cuts between light and dark backdrops and blocking that occupy the entire surface of the frame, demonstrating a high level of planning. This music video also contains five and a half-second long pan, emphasizing a 'domino effect' produced by choreography visible at two minutes and three seconds of running time, the third screenshot from the left in figure 60. Longer pans are a very rare occurrence in vertical videos since the short horizontal axis commonly compromises broad camera movements. However, this video is a testament that longer horizontal camera motions are possible in vertical video with adequate planning. Despite differences in the number of scenes, the two versions of *How Do You Sleep Tonight* share the strongest sense of spatial unity I identified in any example, anchored by the use of similar actors, backdrop, lighting design, choreography, and ending gesture of a single teardrop wipe. Such similarities unambiguously denote the material for both music videos was recorded on the same set, but the vertical version was published following the aforementioned delay of three to four weeks. As a result, the number of views corresponds with the observed trend, where the widescreen version got 300, and the vertical 27 million views, or 9.5 percent of the first release.

The following example, *Chun-Li* (2018) by Nicki Minaj, displays a hybrid strategy for publishing dates and features an aesthetical approach that fit between *Taylor Swifts Delicate* (2018) and *Billie Eilish - You Should See Me in a Crown* (2018), discussed above. The official vertical video for *Chun-Li* was released in April 2018, three weeks before the official widescreen version, (figure 61).

Figure 61: A selection of screenshots from vertical version of Chun-Li released on April 14th 2018.

In this case, Nicki Minaj used vertical video featuring low-quality footage as a teaser for a high-budget widescreen release. To further convey the feeling of user-generated filming by a solo operator, the video contains many shakes and bumps, as well as the frame corrections commonly associated with pre-roll moments when a person is taking a selfie, but on the other hand, the variety of camera angles and tracking shots of a star indicate a presence of at least another camera operator if not the entire production crew. In contrast, the widescreen version of *Chun-Li* (2018), includes a highly choreographed and extravagant version of the track suggesting an early vertical release was a warm-up for the mobile audience and a teaser for users preferring wide and larger screens, (figure 62). In this instance, the vertical video garnered 31, and the widescreen version 150 millions views. The ability of *Chun-Li* (2018) vertical music video to reach 25 percent viewing numbers of widescreen version speaks to the popularity levels of vertical content, but also emphasizes the importance of the first release.

Figure 62: A selection of screenshots from vertical version of Chun-Li released on May 5th 2018.

The last title included in this part of the analysis is *Halsey - Without Me* (2018). *Without Me* widescreen music video, tells a story of two alcoholics breaking up by alternating the couple's scenes of love and violence. The 2.35:1 cinematic version of this text is 30 seconds longer than the vertical video version because of the intro sequence preceding the song's start, (figure 63).

Figure 63: Selection of various scenes from widescreen version of Halsey - Without Me (2018)

In contrast, the vertical version doesn't share any common mise-en-scene elements visible in the widescreen, such as cinematic lighting or spatial unity with any of the scenes, and the editing rhythm is built with a series of M, MCU, and CU shots filmed at a single location, (figure 64).

Figure 64: Selection of various scenes from vertical version of Halsey - Without Me (2018).

As is evident in several of the examples discussed in this chapter, the viewing numbers for widescreen music videos are still far ahead in comparison to vertical video counterparts. However, if I consider the amount of time and affordability of vertical videos featuring a simple mise-en-scene, a re-engagement boost comes as a good return on investment. In the case of Halsey, a vertical version made a difference from losing the Billboard chart #1 slot and going back up in a matter of days. After earning her first solo No. 1 on the Billboard Hot 100 on January 12th, 2019 for *Without Me,* Halsey was dethroned the following week by Post Malone and Rae Sremmurd's track *Sunflower - Spider-Man*. But, on January 26th, she rebounded to the top thanks to a remix with Juice WRLD and a vertical music video optimised for watching on mobile. Released to YouTube on January 19th, 'the video propelled a surge in streaming that brought Halsey back to Billboard's Hot No. 1' (Cirisano, 2019).

In summary, the analysis of posting dates shows the dominance of three to four weeks delay between widescreen and vertical versions in 90 percent of the cases, and in following that trend, vertical videos, on average, acquired ten percent of widescreen viewing numbers. As further assessment of the viewing numbers shows, all titles that made the exceptions and waited much longer than three to four weeks to release the second version exhibited significantly lower audience engagement. Equally important, if the

vertical version was posted first, as in the case of the two Billie Eilish tracks and Nicki Minaj - *Chun Li*, the viewing numbers were comparable to any other widescreen video suggesting the novelty of the track was more important than the selection of aspect ratio. In that sense, any version that is released second, wide or vertical, serves as a booster to re-invigorate dropping audience engagement and music chart ratings. Lastly, having a simpler mise-en-scene the vertical videos often carry the lower sense of production value, making them less attractive for celebrity appearances and product placement, but in return they are much cheaper to produce and convey a sense of intimacy. With that in mind, it becomes clearer why the vertical version is not often the leading one despite the fact it might be a better fit for a mobile device.

Chapter Four: Dissecting Netflix Vertical Trailers

Having an engaging trailer is equally essential in the film industry, as it is necessary to have an attractive video in the music industry since both serve as promotional tool. However, unlike artists, OTT content providers don't own all the movies and TV shows they distribute but rather lease titles for a limited amount of time. Despite having a dissimilar business model to music artists, movie studios and theatres, OTT providers still need to engage the audience and advertise the titles available on the platform, and that is where the reach of mobile media comes into play. Encompassing a very large catalogue, filtered by geographical region, Netflix as a company had to find a quick and affordable solution to produce original vertical trailers for all of its media, suitable for dissemination on mobile devices. Contrasting music videos, where artists have an incentive to produce a new original vertical version of a music video, for streaming services, it is much more efficient to re-use already available footage. Therefore, all Netflix vertical previews are cut from the widescreen trailers with no original material in them. So how does cinematic footage, often filmed in aspect ratios even wider than 16:9, communicate its intended message to the audience in the very slim 1:2 frame from a mobile screen?

Being tasked to make 30-second long previews from a widescreen trailer commonly running for 90-120 seconds, Netflix editors need to re-edit already condensed sequences into an even shorter time frame. For that reason, one common trait of Netflix previews is a leading line of dialogue that starts within the first seconds of opening shots. The spoken words are commonly perceived as voice over because the person delivering them isn't immediately visible, but after several cuts the actual source, often one of the leading protagonist is commonly revealed in the diegetic space, like you can see in the trailer for *The Ballad of Buster Scruggs* (2018) or *Mowgli: Legend of the Jungle* (2018). Another challenge Netflix needed to solve was the fact that the mobile audience might watch previews without sound, for that reason the platform opted to include closed captions over subtitles. Subtitles are superimposed on the image when the audience cannot understand the language of the spoken dialogue, but since closed captioning also includes the descriptions of music and prominent audio FX, they are more appropriate for watching without any audio track. The necessity of including descriptions and spoken dialogue meant closed captioning has a larger character count than subtitles. The next question that comes to mind is why is the number of on-screen characters important for this research?

In a very slim aspect ratio of 1:2, a short horizontal axis limits not only perceived characteristics of the image, but also the number of words that fit per line on the screen. Closed captions and subtitles were traditionally placed at the bottom of the screen in theatres, and they commonly have up to two lines of text, white colour and they never start or end at the very edge of the screen (BBC, 2021, p25). This convention is non-existent in Netflix's vertical previews as text can be found at the top and bottom section of the screen, with words starting only a few pixels away from the screen edge. The letters are coloured grey if positioned at the bottom of the screen or white if they appear at the top. Moreover, depending on the number of on-screen characters, the font size changes accordingly, so fewer words become larger and longer sentences smaller. For example, in some of the analysed texts, closed captions had as many as 18 words and the total of 64 characters on a single title card lasting two seconds as you can see at one minute and 20 seconds of running time in trailer for *Glow Up* (figure 65, second image from the right).

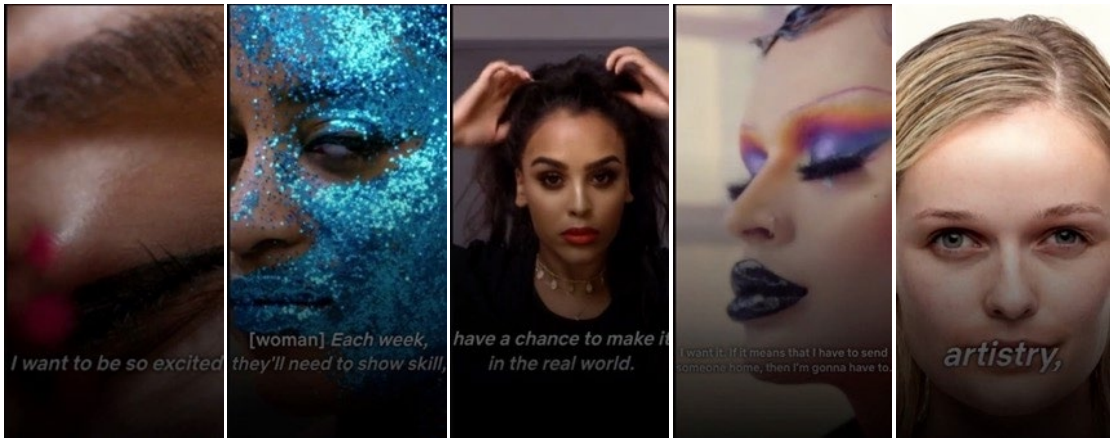

Figure 65: Selection of screenshots from vertical version of Glow Up - Britain's Next Make-Up Star.

Similarly, like any other on-screen element, superimposed text affects the composition of the shot with its position, size and colour, and akin to any unbalanced mise-en-scene, having a text very close to the edge of the screen will produce a sensation of strong edge pull and bring subtext into the sequence. I.e. Having a layer of text in front of the image is a visual distraction since the spectators attention is divided between the action in diegetic space and words, but edge to edge text distribution also conveys the feeling of its continuation in the off-screen space, consequently affecting the perception of frame closure and bringing another layer of connotative meaning into a composition. Lastly, the shifts in text size and position on the screen create an additional visual sensation for the audience, further drawing the attention to itself instead of a principal footage.

Figure 66: Same composition with and without closed captions and gradient mask.

To alleviate distractions made by changing font size, the lower third of Netflix vertical previews has a gradient layer that darkened the image, as you could see in the figure 66. An overlay mask with closed captions is superimposed but not rendered with the video file. Having a separate text overlay means a similar video file can be streamed at multiple geographic locations, while particular language is included as needed. As a primary point of interest, this analysis focuses on the amendments in aspect ratio. Still, because gradient masks modified the look of the principal image, and closed captions influenced its composition, I had to consider how their presence further affected the audience's perception of Netflix vertical previews after an already intricate process of cropping.

The trailer for *Mowgli: Legend of the Jungle* (2018) is included first, as an excellent example of the type of compromise that an editor needs to consider when the wide 2.39:1 footage is re-purposed to the very slim 1:2 composition. As a reference, the shape of a vertical 1:2 Netflix preview frame can fit 4.5 times into the 2.39:1 aspect ratio of the theatrical projection, which means the editor has to cut approximately 70% of the screening canvas. The cutting of such a large portion of the image, without a doubt, affected the perception of the original composition, and the consequences are visible in the scene where baby Mowgli is brought to the wolf pack (figure 67). The first words the audience can hear at the beginning of the vertical trailer are, 'I've never seen a man-cub before; it's so smooth'. The voice belongs to an elderly wolf present in the scene, and visible in widescreen trailer, but after cropping, his body is omitted

Figure 67: Mowgli- Legend of the Jungle (2018).

from the vertical version of the trailer. Because the audience cannot see him in a vertical version, the words of the elderly wolf were perceived as a voice-over, consequently transforming the four-character scene (there is another wolf cub in the original composition) into an intimate moment between a female wolf and a 'man cub', the only two characters visible on the screen. As Smith confirms in his study of *Attentional Theory of Cinematic Continuity*, film trailers command a strong degree of attention synchrony as the gaze stays glued to the centre of the frame most of the time (2011). For that reason, most of the action in vertical trailers also stayed in the centre of the composition, and if there were camera movements, they were mainly executed along the Z axis.

Figure 68: Mowgli- Legend of the Jungle (2018).

Another scene from the film that looks significantly different in the vertical shot is the charge of an elephant toward a hunter. In the widescreen, we see two antagonists, one charging and another standing ground in the same shot, but in a vertical composition, the elephant occupies the entire screen space, and additional cut-away shots are needed to show to the audience the actions of the hunter, (figure 68).

Figure 69: Mowgli- Legend of the Jungle (2018).

While the scene having two contrasting antagonists can be resolved with an insert shot, a dialogue exchange visible at the 21st second of running time proves to be more complex for cropping (figure 69). The bodies of the panther and bear could not fit in the frame together, so they were split in the middle, leaving one visible eye for each character in a very unbalanced and ambiguous composition. Furthermore, as you can see from figure 69, the original widescreen version included five animals with space on the lateral sides for two more. Such blocking embodied unity, strengthened by the stable shape of an arc formed by their bodies in the composition. I would argue that a scene like this cannot be presented in a single slim shot at all, and unlike the scene with a baby, where the meaning was altered but the characteristics of relationships between the protagonists stayed the same this scene lost most of its connotations due to cropping.

Figure 70: Mowgli- Legend of the Jungle (2018).

The last two screenshot pairs, figure 70 and figure 71, show how cropping altered the perception of volume duality in the shot but did not change its intended meaning. Even after losing the forest on the left and the mountain range on the right from the screen, the audience would still have enough visible elements to create a desired meaning in the diegetic space, a rider, distant mountains, and an open sky (figure 70). In this case, that works as the scene depicts cheerful Mowgli touring the jungle on the back of his black panther friend Bagheera.

Figure 71: Mowgli- Legend of the Jungle (2018).

Similarly, despite cutting the dominant positive volume of a bear's body from the frame, the intensity of the apes' jump wasn't diminished, (figure 71). Even though there was no visible target in the slimmer shot, the ferocity of the attack was intensified because the ape body occupied the entire frame, producing the sensation there was no space to run as it moved swiftly towards the audience across the Z-axis. These findings were similar to observations in vertical music videos that scenes with fewer or single protagonists benefited from tight composition, but as you can see from the following

example, *Formula 1: Drive to Survive* (2019), the framing of an object such as a Formula 1 car on a race track, pose a different challenge for 1:2 aspect ratio.

Because of the vehicle's shape and slim horizontal profile, a vertical shot fully encompassing a wide object like an F1 car will include an unusual amount of headroom. Furthermore, if the composition needs to include multiple objects with similar characteristics, such as two rows of cars lined up at the start, and another even wider structure such as a pit lane gantry, the distance necessary to encompass them all may affect the number of details the audience can perceive as you could see from the last slide in figure 72.

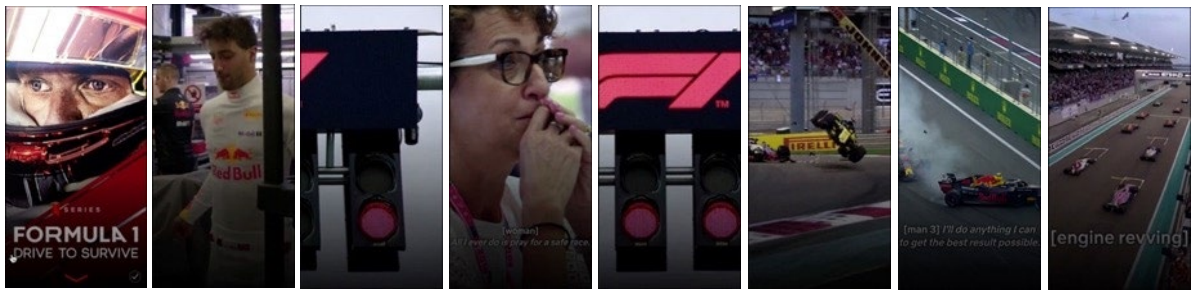

Figure 72: A selection of screenshots from the vertical trailer for Formula 1 - Drive to Survive Netflix show.

The signalling unit above the start grid contains five double rows of round lights, that once illuminated in unison and turned off, mark the start of the race. The shape and size of the gantry pose a compositional challenge for a slim 1:2 aspect ratio, but because it features an important story-driving element, it has to be included in the sequence. For that reason, the editors of the *Formula 1: Drive to Survive* (2019) vertical preview divided a single widescreen shot of the starting lights into three vertical parts and, through the use of montage, convert it into an object that drives the tension for the first half of the trailer's running time. If compared to the widescreen original, (figure 73), by not including the entire structure in a W shot, the editors emphasised the hermetic characteristics of a vertical composition and applied them to build suspense with a series of M and CU shots. The described changes are a prime example of skilful editing, since the sequence preserved the story dynamics and context of a suspenseful F1 race start, despite drastically modifying the look of the original footage.

Figure 73: Widescreen screenshots from Formula 1 - Drive to Survive show.

Figure 74: Formula 1 - Drive to Survive.

With that conclusion in mind, in the following few paragraphs, I will elaborate on how reframing the remaining key scenes in the *Formula 1: Drive to Survive* (2019) vertical trailer affected the overall perception of the presented visual sequence. From figure 74, you can see how the aforementioned black gradient filter produced an image that appears more 'high contrast' than the original. The bright spot on the helmet visor became juxtaposed with an almost complete black lower portion of the shots, taking the audience's attention away from the drivers' eyes. In combination with a tight 1:2 aspect ratio, lower luminance levels of the shot turned the already hermetic CU of Daniel Ricardo concentrating before the race start into an ECU shot, where visibility

of only one eye supports the feeling of constrained space and produces a more sinister interpretation of his thoughts before the race.

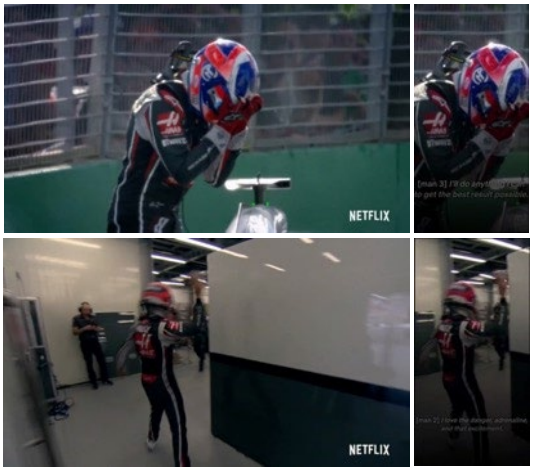

Figure 75: Formula 1- Drive to Survive.

Following a similar logic, the reactions of Hass driver Romain Grosjean after his crash are also perceived differently concerning the composition and framing used to present them to the audience, (figure 75). In the M shot of Grosjean's reaction after the crash at the side of the racetrack, his gesture signals despair, but at the same time, he also shares the space with his car, the large positive volume of the security fence and fans behind it. In contrast, in a cropped vertical version of the exact moment, Grosjean looks more isolated and alone with his thoughts due to the lack of lateral space in the composition. Additionally, once merged with closed captions and a gradient mask, Grosjean's downward movement appears to land in a much darker negative space than the lower third of the original widescreen composition occupied by the bright silver tip of his F1 car. Once the documentary coverage continues inside the pit, his angry throw of gloves can also be coded differently. As the camera tracks behind the agitated driver, a widescreen composition includes a member of the Hass team using a phone in a position like he is sending a text message. As soon as Grosjean hurtles the glove, and the audience cannot see where the glove lands due to a quick montage, the entire moment renders a feeling of an irritated driver targeting that particular member of the Hass team for whatever might be his role in the crash. Consider a similar action in a cropped 1:2 aspect ratio, where the body of a fast walking, angry driver occupies most of the composition. In that case, a person sending a text message is significantly less prominent, so Grosjean's glove thrown into off-screen space now appears without the target, so is arguably more self-directed. No matter how brief and subtle, such changes are perceived subconsciously by the audience, and they will affect their judgment of the event; in this case, if the driver is frustrated with himself or the team.


Figure 76: Marriage Story (2019).

Editors working on the trailer for *Marriage Story* (2019), a feature-length drama about a couple separating, faced a similar creative challenge. The topic of the movie may not bring to mind a fast editing tempo, but the vertical trailer of this film has 36 shots in 28 seconds, for an average of 0.93 seconds per shot, which is just ten percent slower than the pace of the trailer with Formula 1 races as the subject. With a running time of 28 seconds, *Marriage Story* (2019) is the shortest vertical preview I dissected in this research, but regardless of its brief length, this text is prolific with dissimilar connotative meanings instigated by reframing. As you can see in figure 76, a courtroom establishing shot was designed to have strong overhead converging lines, leading to the judge's bench as a decisive figure in the composition. Two confronted groups are separated by the judge's authority in the middle of the shot, conveying the sense that both parties are equal in the oncoming trial. Once the defendants and legal assistants are cut from the group of three, the newly formed composition becomes a 'battle' between the two representatives. Additionally, without converging ceiling lines and with the introduction of a gradient mask, the persona of the judge is not only presented as stronger by having more screen space and a higher position, but the power of his authority becomes more decisive as his bench marks the edge between the bright/high key part of the frame above him and the low-key section occupied by rest of the people in the courtroom.

The following screenshot from a scene where the father reads a bedtime story to the child (figure 77), the widescreen composition includes kids' toys stuffed in the bed above the pillows of the troubled family and is overall brighter. Without the additional space visible in the widescreen, the trio appears tucked closer together in the 1:2 frame, with the face of the mother hidden under the gradient filter that obscures her tears for the audience.

Figure 77: Marriage Story (2019).

Figure 78: Marriage Story (2019).

Figure 78 also demonstrates how similar words and blocking can carry different weights depending on the framing. As the father expresses his frustration to his lawyer in the widescreen shot, the index vector and motion vector produced by his body position and moving hand lead the audience to gaze to the right while leaving negative space on the left-hand side. Ample empty space in widescreen composition signifies his 'missing family,' while the proximity of a corner suggests the lack of progress in his legal case. Once the same monologue was delivered in a vertical shot, the composition did not include the negative volume of the empty wall or the connotative meaning of the person's situated close to a corner. Housed in a vertical medium shot without visual elements to challenge his presence in the composition, the words the father delivers, as well as his attitude, would appear more decisive to the audience.

The last set of screenshots from Marriage Story (2019) (figure 79) summarises how influential the graphical user interface[x] (GUI) can be on the aesthetics of the shot, since presence of an overlay mask carrying text modifies the image luminance values and alter the perception of colour in the composition. In one of the final scenes, as parents passionately confront and remind each other of the past, a chance for positive resolution will be perceived differently by a mobile audience and spectators watching this drama on a larger screen. In the W shot, the mother and father are alone in the room with no specific visual features, as uniform colours and the lack of personal details embody their relationship's current state. Nevertheless, while the room's look suggests emptiness, its size hints at options for the couple, and higher luminance levels support the impression of a positive end. In contrast, if I compare the atmosphere between two vertical screenshots in figure 79, one with and the other without a gradient mask, the vertical 1:2 composition brings a feeling of a more confined space where the couple has fewer options, but they are still focused on each other. Knowing the plot of the film, that connotation is still welcomed, but it is once again a gradient mask that alters the meaning more than a frame crop. A vertical frame with a closed caption mask renders a 'darker' mood, as the figure of a kneeling father is largely obscured and lower luminance levels of the scene suggest 'less promising' positive resolution due to its darker tone.

Figure 79: Marriage Story (2019). Comparison of original widescreen, vertical shot with and without gradient mask.

The following sample, *Inside Bill's Brain: Decoding Bill Gates* (2019), shows that even in a condensed editing sequence housed in a vertical frame, there is a space for camera pans, tilts, and dolly movements, (figure 80).

Figure 80: Inside Bill's Brain: Decoding Bill Gates (2019).

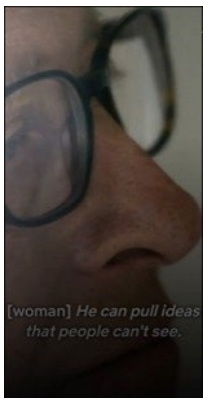

Figure 81: ECU with minimal lead space.

Although all observed secondary motions were brief, they were highly motivated by the trailer narrative by crafting the feeling that 'something is always moving' inside the protagonist's brain. Furthermore, since the leading character was often situated inside secondary framing or behind other positive volumes, such compositions further alluded that in order to achieve his goals, Bill Gates must tackle multiple obstacles. At the 13th second mark, I spotted a shot featuring a direct address method (a middle screenshot in figure 80), and unlike music videos where this technique is used often, a single direct gaze into the camera makes this gesture stand out as a strong accent in the cinematic sequence. Lastly, this trailer also demonstrates how minimal lead space for index vectors, such as the protagonist's eye line in figure 81, is sufficient to project a clear frame intention even in the slim composition of a 1:2 aspect ratio.

The last screenshots I have included in this chapter were taken from Coen's brothers *The Ballad of Buster Scruggs* (2018). From figures 82 and 83, you can see despite the great variety in shot scale, none of the depicted compositions lost any of their principal connotative meaning despite cropping and closed captions masks. A desperado that hesitantly walks into the bank and gets greeted by a quirky teller is still boxed in a secondary frame produced by the shape of the door (link to vertical preview), hinting at a potentially bad outcome of the robbery, and the banker's reaction is clearly visible in the slim frame, (figure 82).

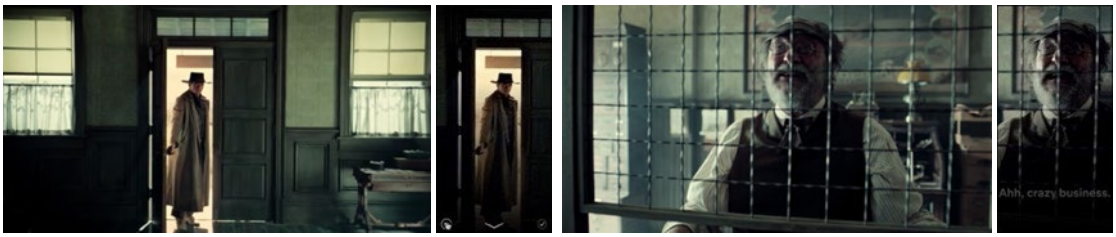

Figure 82: The Ballad of Buster Scruggs (2018).

Similarly, in figure 83, a W shot of a miner admiring the valley he just discovered is understood in the same way, despite the fact the audience lost the context of the forest he just walked out of, shown by the tree on the right and branches on the left of the original frame. Lastly, a lack of worn-out wall in the background, and belt buckle in the foreground, did not change the meaning that a cowboy is preparing to draw a gun in an imminent duel. The rest of the footage used in the editing of this trailer fits into a similar pattern, and a few rationales could explain why cropping was relatively simple.

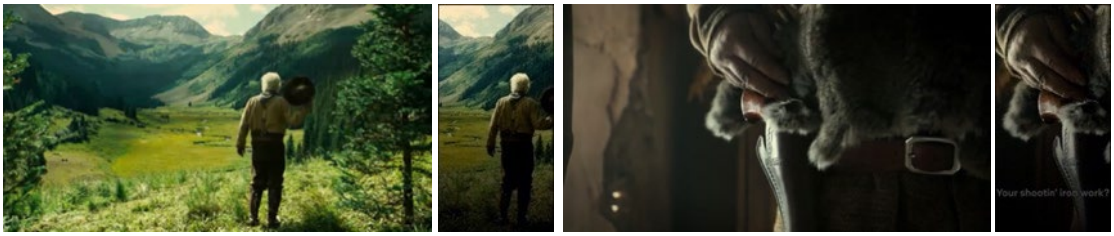

Figure 83: The Ballad of Buster Scruggs (2018).

Firstly, the well-defined compositions commonly featuring a single protagonist in a wide-open countryside come as one of the Western genre conventions. Second, the scenes of gun duels frequently featured in *The Ballad of Buster Scruggs* (2018) should be simpler to re-cut since the protagonists are in the same space but never close together, as they might be in the dialogue scene. For that reason, the use of a narrow vertical 1:2 aspect ratio did not change the meaning of the scenes. Furthermore, comparable to vertical music videos, sequences where subjects move directly towards the camera along the Z-axis benefited from narrow framing since the forward motion was intensified, but some of the elongated 2.35:1 compositions, such as multi character ones seen in Mowgli, have undergone a transformation that significantly affected the audience perception of the original scenes.

The editing decisions observed in the rest of the key texts sampled from Netflix followed the already presented pattern. For that reason, I did not include any screenshots from samples #6 *Orange Is the New Black*, #7 *Patriot Act,* #8 *Stranger Things,* #9 *Sugar Rush*, and #11 *White Fang* in this chapter.

Chapter four: Tall Music Videos & 1:2 Film Trailers – Conclusion

The most notable new-found attribute relevant to this research was a substantial difference in complexity of mise-en-scene between the official widescreen versions of music videos and their tall counterparts. Furthermore, several A-list artists' single take vertical music videos featured an unusual repetitive camera movement, aimed to increase the levels of audience attention synchrony due to a lack of edit points. A detailed breakdown of shots used in key music video samples revealed a minimal difference in the number of medium compositions between widescreen and vertical texts, as well as the complete absence of vista shots. Vertical videos feature 11.61 percent more CU and ECU shots, while widescreen had almost ten percent more W shots, but an even more apparent difference between the two was expressed in significantly fewer edit points found in vertical texts. The editors of key music videos used 1174 shots to build 11 widescreen sequences, in contrast to only 512 shots found in vertical versions, meaning vertical videos had 57 percent fewer cuts. Five single-take music videos contributed to the large difference in overall statistics, but even when I compared titles individually, vertical versions always had between 25-65 percent fewer edit points.

Analysis of 2D and 3D characteristics confirmed earlier observations that the attributes of human proportions leave limited space in vertical shots for other two and three-dimensional structuring elements. A single performer filmed in an M shot and housed in the middle of a vertical frame creates an exceptionally balanced composition due to the central distribution of dominant graphic mass, and equal distance of a principal subject from all four edges. For that reason, the compositions in vertical videos contained less secondary framing, two shots, over-the-shoulder angles and explored less storytelling potential of volume duality than widescreen music videos. A greater sense of intimacy explains the popularity of vertical videos with solo artists and its appeal to fans, but at the same time, it contributes to the rationale for using a mise-en-scene that lowers the production cost. I would also argue that a simpler mise-en-scene is a principal reason why I have not identified any product placement in vertical videos, since reduced details in composition offer less chance for subtle advertising that will not turn a music video into a commercial. Although lower production cost is commonly welcome, the majority of A-list artists still preferred high budget widescreen productions for the premiere of their new tracks and commonly used the vertical version as an engagement booster for a mobile audience three to four weeks later.

In contrast to planning associated with the production of music videos, the editors of Netflix previews did not have a chance to work with the original footage, and that limitation compelled them to make significant creative compromises during the cropping process that, in return, often altered perception of compositions' stability and characters' relationships. The lack of composition balance and clear meaning caused by cropping occasionally exhibited in vertical footage, was partly alleviated by brief running times, indicating vertical trailers can compensate for perceived lack of stability better than music videos or other longer forms, where the audience has more time to analyse the frame in greater depth. Despite brief running time, Netflix vertical trailers used a greater variety in shot scale than music videos,

including vista shots, but due to the significant difference in the shape of the frame, scenes featured in them experienced substantial transformations. I would argue the editors utilised vista shots only when there were no other means to establish characters' location, as I identified only two in 352 shots. Such low numbers are testament to the limitation of the extremely elongated 1:2 aspect ratio to house very wide compositions, and the scale of editors task becomes even more apparent when I take into account that most of the original widescreen material for trailers was filmed in aspect ratio wider than 16:9. The comparison of average shot length between the original and cropped version did not display any significant patterns, suggesting the shape of the viewing canvas did not play a major role in the determination of trailers editing pace.

In some cases, like in the example of *Formula 1 - Drive to Survive*, skilful manipulation of the footage in post-production preserved the scenes' emotions, stayed true to the original atmosphere, and arguably introduced a new layer of welcomed emotional subtext. However, fast editing couldn't always hide the limitations of a short horizontal axis or problems visible in the dialogue scene from *Mowgli*, when the stable composition of five protagonists was transformed into ambiguous shots of two. Comparable to the substantial influence of cropping on the perception of the scene, it was not possible to ignore the effects of a gradient mask and how the inconsistent style of closed captions influenced the perception of the footage. Lower luminance levels often introduced a 'darker' tone in the composition, but even more important, the presence of a mask sometimes obscured subtle story driving elements such as tears on Scarlett Johannsen's face in *Marriage Story*.

In conclusion, in vertical music videos and vertical cinematic sequences, the screen energy of an object moving fast along the Z-axis was enhanced due to a tighter frame, and both genres used that fact to create visual accents. The trade-off for benefits offered by a more defined shot was visible through the protagonists' lateral motion limitations and a lack of elaborate camera movements. Regardless of the original widescreen aspect ratio of a trailer, which ranged from 16:9 to 2.35:1, the compositions with multiple mise-en-scene elements of equal importance were the ones that exhibited the most extensive transformation in the process of cropping. Compositions featuring a single protagonist or point of interest were the simplest to re-frame, and they occasionally gained an additional layer of welcomed subtext once housed in a 1:2 aspect ratio. Nevertheless, the majority of re-purposed footage came across as a less effective alternative against the full expressive potential of the original, which highlights the importance of advanced planning and thinking of aspect ratio in the early stages of production. Lastly, these findings were beneficial during the production of thesis artefacts, when the aesthetics of vertical video, both original and re-purposed, were tested experientially in chapters six and seven.

Chapter Five: Talking to Professionals – Introduction

The following step in the research led to conversations with media professionals, who produce content for a range of legacy and online distribution platforms in an assortment of geographical locations. By reaching out to working professionals, I sought to confirm or dispute observations provoked by the systematic textual analysis up to this point. Recognising I am studying an expressive practice not fully adopted by the film and television industry, I also wanted to learn more about technical and aesthetic decisions professionals make when they produce vertical content. My use of semi-structured interviews and open-ended questions proved to be a good choice for the topic I intended to cover, as it left abundant opportunities for exploring diverse personal perspectives and insights.

In an effort to cover various skills and media markets, I identified six participants on three continents. The youngest interviewee Anna works for the TVNZ online program division in Wellington. In her production environment, Anna has significant creative freedom regarding the choice of topics and the aesthetics of visuals since a segments her group produces doesn't have to follow traditional TV program. At the same time, working for the largest public broadcaster in any country points towards a structured environment, where key production decisions need to be approved by multiple tiers of senior management. Elsie also works in the New Zealand market but is based in Auckland and employed as an editor in the production of high-budget TV shows for global streaming services such as Amazon Prime and Apple TV. Max, based in Australia, is one of the representatives of the mid-generation. Max's inclination towards auto ethnographic and documentary research makes him a very open-minded creator, a cineaste always on the lookout for new ways of storytelling and visual content distribution.

For the fourth participant Roi, a freelancer from Spain with a background in directing theatre performances, cinematography, and editing, the choice of a recording device, aspect ratio, or any other image parameters comes down to their advantages for a particular production. He deems any video format or aspect ratio as good if it can progress the story and improve the audience immersion in his transmedia projects. Like Elsie, Claire works on high-budget TV shows produced for streaming services, but she is based in a larger United States market, traveling between Los Angeles and New York City. In the role of position of director, Claire is a decision-maker responsible for key aesthetic and technical judgments, making her insights and comments on specialised vertical camera passes in high-budget production invaluable for this research. Lastly, with over 30 years of experience as a producer/host of news programs and magazine shows, Ernabel embodies an older generation of TV professionals whose portfolio pre-dates the digital revolution of the 1990s

In the acquired interview data, I identified several common themes. The first one was a strong personal preference for widescreen aesthetics, and echoes of this premise were recognised in most of the data. The second theme was the most complex to isolate, since the discussion included a few overlapping technical and aesthetic facets of vertical video. All participants agreed it is necessary to develop platform-specific content for the optimal performance of videos, but at the same time, there was a division of opinion regarding the acceptable length of vertical content and affordances of the short horizontal axis. The third theme was the division of opinions if smartphone mobility is an acceptable trade-off for lower footage quality and means to create aesthetic variations in the footage. The last major theme was the impact of the COVID 19 pandemic on the media industry, linked to the fact there is an abundance of vertical content in contemporary social media, so despite a personal opinion or aesthetic preference, interviewees acknowledge if the audience accepts the format professionals cannot ignore it.

While some of my methods may be analogous to production culture studies, the rationale not to engage with critical questions related to production cultures primarily came from my desire to conduct practice-led research and to explore practitioners' personal choices and preferences related to the vertical format. During the interviews with practitioners, I wanted to focus on technical and aesthetic decisions related to the affordances of vertical content but not to explore the broader context of the participant's work

environment. Therefore, I did not discuss topics such as budgets or other possible restrictions related to a particular company, nor did I pose any questions regarding company traditions. Furthermore, my research's timing also influenced some significant limitations on the prospects of such a study. Due to various global restrictions imposed by COVID-19 lockdowns during 2020-2021, I could not be present in a work environment with media professionals to study their production workflow, observe their immediate collaborators and assess the influence of supervisors on their technical and creative decisions. Lastly, due to travel restrictions, I would have needed to limit my research scope to the territory of New Zealand, which would significantly affect the diversity of my data sample. In contrast, online interviews and a research methodology that did not require fieldwork to observe practitioners' work environments made it possible to conduct the study during the travel restrictions of the pandemic and enabled a diverse sample of interview participants.

Chapter Five: Theme 1 - Strong Personal Preference for Widescreen Aesthetics

As the technology progresses and offers new simplified video recording workflows, specialised roles such as producer, cameraman, sound recordist and editor, gradually blend into a single person under the term media content creator. While embracing the versatility of new roles, contemporary content creators look for the tool to fit their multitasking abilities, and they found the match in modern handsets. There is arguably no other device as versatile as the smartphone regarding its recording capabilities, form factor, built-in options for post-production, and the ability to immediately distribute recorded content online. All interview participants were aware of technological breakthroughs associated with smartphones, and some already included them in their workflow. Still, the comments were much stronger and, in some instances, very passionate during the discussion on changes vertical aesthetics brought into online video aesthetics. But what happens when the audience is more flexible and willing to accept a new format than professionals favouring widescreen aesthetics as in the case of the vertical video?

These tensions between new and well established formats supported by strong personal preference were evident in the comments of Elsie, Claire, and Roi, as they deemed a vertical projection canvas of an upright smartphone unsuited for narrative and cinematic content. They used attributes such as ‘cinematic’ and ‘beautiful’ to describe widescreen while arguing that the true emotion and depth of a story cannot be conveyed through a slim canvas or small mobile screen.

> It's simply impossible to show the same thing vertically as you want to horizontally. If you're watching a narrative, and you want to sit down and get into a story, the odds are you're not going to watch it on your phone; you're going to watch it on a television set, or a computer screen or even an iPad (Claire).

Claire's inclusion of an iPad in the range of screens suitable for a cinematic narrative is intriguing because it corresponds with findings presented earlier in this research, chapter two/figure 7, the influence of screen size diagonal on a smartphone and tablet orientation. Scientia Mobile found that 93.9 percent of the audience holds their mobile device horizontally if the screen diagonal is 11" or more. That also indicates that Claire would consider any widescreen canvas, including an 11" mobile screen, but not vertical orientation for cinematic content, highlighting that her antagonism toward vertical content is more related to screen orientation than the screen size.

Likewise, without getting into details of her preferred canvas for cinematic content, Ernabel stated she never thought of vertical video as a unique format before preparing for her interview with me.

> Until you brought this up, and I was reading your questions, I had never thought about the vertical format as a separate entity. Maybe because I'm old school and I've been doing this for a long time, I haven't made that transition. I'm still thinking as a widescreen videographer, and I don't think my stories will look good in the vertical orientation (Ernabel).

Elaborating further on her role as a host and a producer interested in expanding the reach of her show, Ernabel stated she perceives the vertical aspect ratio as a supplementary format, an accompanying

promotional version of original TV segments aimed at social media audience. A similar perception of complementary but not a principal piece, was prevalent in the relationship between widescreen and vertical music videos analysed in chapter four. Claire's push back goes even further, since she as a director thinks that vertical composition compels her to make aesthetic compromises by including irrelevant information in the composition that does not align with the human field of view. 'It's very hard to tell narrative stories vertically; it's just not the right format. Another problem I have with vertical is not how eyes see, we don't see vertically' (Claire).

Looking back at my professional television and film career before commencing this research, I can verify the majority of my colleagues would use similar words to describe the restrictions of a vertical frame. But when Claire reflected on her experience when filming for Quibi, she pointed out more specific creative compromises caused by the slim filming frame, most notable the inclusion of unnecessary information in the composition. 'If I'm looking at you, I'm not looking at your feet, I'm looking at your face, but with Quibi shot, the frame encompasses all of your body. It's boring information' (Claire). From her perspective as an editor, Elsie depicts a similar sentiment towards vertical content as she poses a question.

> Why would you shoot vertically in the first place? To me, it's restrictive in terms of aesthetics. If you have a limited screen time and need to make a choice, why not go for this way [gesturing widescreen] and give the audience more (Elsie).

A personal preference for widescreen was commonly intertwined with a discussion on the technical parameters and the supreme quality of footage associated with larger cameras. Elsie indicated her trained eye could immediately recognise the difference in footage quality even if the shots were intercut in a fast-paced action sequence. As an example, she recalled her experience on the set of the Lord of the Rings and Hobbit trilogy.

> I remember they were shooting on RED 4K or 5K at the time, together with some GoPros. When you watch it on a cinema screen, it doesn't look as good, and at least to me, GoPros footage stands out like a sore thumb (Elsie).

It is understandable some professionals are more resistant to the changes in traditional workflows and conventions of the language of moving images because amendments to the well-established methods often bring a disruption. As Manovich summarised in *The Language of New Media*; 'with new media, a new area has emerged. As the 'professional' technology becomes accessible to amateurs, the new media professionals create new standards, formats, and design expectations to maintain their status' (Manovich, 2001, 133). Nonetheless, the speed of remediation associated with new media urged professional creators to re-evaluate their traditional aesthetic preferences, become less resistant to vertical content and be aware of changes online distribution platforms are bringing into the language of moving images and standard TV/film workflow. The strong preference for widescreen aesthetic meant that most participants automatically associate vertical content with short forms and user-generated content, discarding even consideration of vertical canvas for cinematic content and long forms. A strong antagonism toward vertical aesthetics also led to an inability of some participants to objectively assess the potential benefits and limitations of vertical content. The same objections resulted in a lack of interest and, consequently, effort to research any contemporary professionally produced vertical content that might influence or change their opinion.

Chapter Five: Theme 2 - Platform-Specific Content, Short Horizontal Axis, and Brief Running Time

The expressive potential of the vertical frame was mainly discussed in the context of limited camera movements in cinematic sequences, and the inability of a short horizontal axis to frame multiple people in the same shot as commonly seen in TV sit-down interviews or studio talk shows. The data also revealed interviewees shared a similar opinion when it comes to the need to tailor a script to the aspect ratio, but it was challenging to determine to what extent participants blamed the characteristics of the vertical frame for their creative limitations against the shortcomings of social media platform where

they would like to have an online presence. Equally important, all interviewees except Max associated a shorter horizontal axis exclusively with brief running times. Short running times were further linked to a simplified mise-en-scene, and because of those two factors, the vertical frame was deemed inappropriate for cinematic content by Anna, Elsie, Claire, Ernabel and Roi. Two participants, the youngest (Anna) and most experienced (Ernabel), both stated they struggled to re-work their widescreen stories or produce more original vertical content in the past despite being aware of its engagement potential on social media platforms. As a principal reason for the inability to produce multiple aspect ratio versions of their stories, they blamed the lack of resources in their departments and the limitations of short horizontal axis of a 9:16 aspect ratio as an inadequate frame for many of the scenarios they commonly cover.

> We made custom pre-set widescreen sequences and have a show without the presenter. For that reason, it wasn't only about repositioning guests in the vertical shot. Knowing many people watch videos on silent, and without the option to use voiceover, we rely a lot on subtitles and graphics (Anna).

Recalling the findings from the discussion on Netflix vertical previews in chapter four, I came to a similar conclusion as Anna. Re-purposing is not always a matter of a simple crop or footage blow up, and a shorter horizontal axis, without a doubt, compromises the ability to use longer subtitles sentences. In Ernabel's case, commonly shorter running times of social media cut-down posed an additional challenge to the already complex re-purposing process.

> Our team just posted the same video without worrying about how distorted the video is going to look, and that did not worked well for us. Instagram TV needs vertical video but Instagram Stories didn't. In any case, because the video was too long for Instagram, I was getting cut off in the middle of the sentence before the end of the segment (Ernabel).

Not having the resources to develop a platform specialised content, Ernabel decided to pause the activity of her show on Instagram, and attempted to expand the *Asian American Life* audience by posting on TikTok, but still without fully following the preferred technical specification from the platform.

> We decide to start reworking length of some of our old videos to put them on TikTok, but we didn't change the aspect ratio to vertical. We keep the same widescreen, and because of that the video had a weird look (figure 84), (Ernabel).

Although the length of the video was now appropriate, Ernabel learned that if a segment wasn't created for a particular social media platform on a producer's mind, an inadequate aspect ratio would also significantly affect the quality of the viewing experience for the audience, and potential reach of the video. For that reason, after a few months of using posts re-purposed from widescreen footage and without a budget to produce platform-specific content, Ernabel and her team abandoned TikTok, acknowledging the context of the stories made from cropped widescreen footage was often misleading due to a lack of original widescreen elements, to the point that Asian American Life show not being present of major social media platform is more reasonable instead of posting stories featuring ambiguous context.

Figure 84: By weird look, Ernabel refers to large letterboxes exhibited when 16:9 footage is watched at a 9:16 aspect ratio screen without a blow-up.

Anna was in a similar situation since her department commonly needed to produce a widescreen version of a story for the TVNZ website and two shorter social media cut-downs. Still, her team stopped making vertical versions since many segments featured more than one person, and the crew found the re-purposing process very difficult. 'We make widescreen web version, then a one to one (square frame) cut down followed by vertical. But it was really difficult if you're trying to have a two person interview, that was a nightmare' (Anna). Anna's remark is also a testament that having a script written explicitly for a vertical aspect ratio is equally important in the professional television and film industry. After acknowledging the necessity to follow the platform's technical specifications for a better performance of videos on social media, Ernabel requested an additional budget from her TV

station and employed another video editor. The newest member of *Asian American Life* team was tasked to re-cut the original widescreen segments broadcasted on CUNY TV into shorter versions for social media platforms, but that initiative still did not solve the problems associated with the repurposing of widescreen footage into the slimmer vertical frame.

> We'd give our social media person a lot of editorial decisions on how she cuts down the segment, and our viewers now can see what is going on without it having an awkward zoomed-in look. She's done a good job overall, but still, it would be good to produce more original vertical videos (Ernabel).

This statement also shows that in the industry such as broadcast television, the personal motivation of a producer is just one part of a complex workflow, since the production of unstandardized formats such as vertical video asks for re-evaluation of well-established storytelling methods. It is important to stress, the amendments to production workflow are not always associated with investments since vertical content filming does not require specialized gear. Still, exploring alternative storytelling methods requires active engagement in research, and in fast-paced production environments, it is not always easy to find time for that. Despite their personal preference, Claire and Roi also recognise the necessity to develop platform-specific segments in the free-lance domain of media content creation and think of the appropriate aspect ratio, length and editing style for each distribution outlet.

> It's a matter of creating a video for an application designed for use on a smartphone, in that case I think vertical works. Short, fast and funny, under a minute. If it's much more than that, I just don't see people wanting to stare at their phone for so long (Claire).

The length of time a user spends on average on a mobile device was not on the list of questions. However, after Claire's referencing only very short forms multiple times in her answer, it was clear she does not consider a vertical format and mobile screen appropriate for any video forms longer than trailers. Roi was more open to explore the creative possibilities of vertical video despite its short horizontal axis. For him, adapting footage to the distribution platform meant having an attractive montage in the first five seconds.

> It is very important to make the first seconds of the video strong by including catchy content. This is crucial; otherwise, you're going to lose the people's attention, which is different from cinema, where you can wait 30 seconds or more to get to the point (Roi).

He sees the aesthetic shift from long establishing shots as crucial for the success of videos his production company prepares for YouTube. At the same time, although he accepts vertical video and fast-paced editing as necessary for social media, he often hires another editor for social media cut-downs since that is not his preferred style.

Looking back at the history of moving images, it is clear that the adoption of new formats by media professionals was never a straightforward transition. The debates surrounding the issue of framing for multiple screen shapes is almost a century old. During the 1930s in the US, the Academy of Motion Picture Arts and Sciences saw the situation around standardization of widescreen and cropping issues of new talkies as a serious problem, and as a solution, proposed a special 'Memorandum on the Problem of Aperture Shape' in August 1929. Frank Woods, the secretary of the academy stressed:

> The theatres in the country are projecting talking pictures through apertures of three different shapes. As a consequence of unstandardised filming and projections, the production workflow was compromised since the studio cameraman had to compose for either two or three alternatives of frames size and shape (Hovet, 2017, p.141).

However, how different is a vertical filming workflow for other non-camera departments involved in the production? On the question of the amount of time necessary to plan vertical composition, Roi responded that a vertical frame requires more pre-production time, especially if mise-en-scene incorporates the camera movements and multiple subjects. That also includes set designers and studio crew that had to be mindful of the top sections of the set or even the look of the ceiling since the Y-axis of the frame is much longer in vertical composition than in the widescreen. He also brought up intrinsic perceptual features associated with the geographical location of creators and audiences:

> In Western world, our visual system reads the image from left to right. In contrast, a vertical frame reads up and down, with not much concern over the space of left and right, so if you want to incorporate a movement, you have to plan more (Roi).

Roi continues that cinematic tools for the building of diegetic space are limited in the 9:16 frame due to a lack of space on the horizontal axis. 'Off-screen space is perceived differently. In the traditional cinematic language, we incorporate the off-screen space in a narrative. But in a vertical frame, my mind is still struggling to manage the limited screen space; for me, the horizontal edges are too close' (Roi). This thought correlates with Claire's comments from the position of a director in high-budget production**.** Claire would do a dedicated Quibi camera pass if a vertical frame's composition was compromised by the broad movement or inability to house multiple subjects.

> We would usually run the other two cameras concurrently for widescreen, but sometimes I would do a dedicated Quibi camera pass if it was just an impossible shot. We would repeat the whole scene again, with a focus on the Quibi camera (Claire).

By impossible shot, Claire indicates mise-en-scene that couldn't fit in the vertical frame. A separate vertical pass was necessary for all scenes that left any signs of ambiguity, since the platform gave equal prominence to the widescreen and vertical view of the scene, not knowing in advance what orientation the audience would choose. At the same time, as the camera crew adjusted their position and sometimes changed the lens to adjust the perspective of the shot, Claire's instructions for actors' delivery and blocking during the vertical camera pass stayed the same.

> There was really no difference for me because the words were still going to be the same. The action still happened in the scenes in the same way. The actors still spoke words at the same pace. If someone was affected, that would be the camera crew (Claire).

In summary, the diversification of online distribution outlets, without question, created a new challenge for contemporary media content creators because it is often not possible to use the same footage for widescreen and vertical canvas without aesthetic compromises. Furthermore, as Anna and Ernabel argued, not all departments have the budget to shoot dedicated vertical footage and develop platform-specific scripts like in the case of Quibi productions. Equally important, since the majority of studios were built with widescreen composition on creators mind, a potential shift in aspect ratio means many spaces becomes inadequate for vertical filming as tall frame would reveal the ceiling, unless the workspace has an extremely high ceiling. Considering significant investments already made in studio spaces all over the world the issue of professional vertical production becomes more of a problem of a limited amount of time, and lack of specialised spaces able to accommodate the vertical composition than restrictions caused by a shorter horizontal axis.

Chapter Five: Theme 3 - Mobility and Ergonomics Against Footage Quality

Similarly, like a theme of personal preference, themes of smartphone mobility against footage quality a device can record frequently re-emerge in the conversations, even when questions examined interviewees' opinions on another topic. A block of question on the use of a smartphone as a primary recording device and the debate around the distribution outlets revealed media professionals are acknowledging the improvements in smartphone technology and the necessity to produce platform-specific content. But despite significant improvements in the video recording capabilities of smartphones, media professionals often think the footage quality is still not satisfactory for professional workflows. Moreover, regardless smartphone's touch screen technology being very intuitive for everyday use, similar traits pose a problem when a smartphone is used as a primary recording device in a professional workflow.

Elsie and Claire could not find the rationale for using a smartphone if there is a budget for a professional camera. They argue the trade-off between savings in the camera department was not justified enough since acquired footage is of lower quality, and lack of standard often creates multiple difficulties for the post-production crew. For those reasons, post-production setbacks accumulated in the long run

usually cost more than the initial saving made by the use of a non-professional camera system. Likewise, for Elsie and Claire, the use of a smartphone or other consumer cameras such as GoPro is a limiting factor for creativity since footage recorded by a professional camera body can be manipulated to resemble the look of an amateur recording device, but not vice-versa.

> Instead of using smartphone or GoPro, a production should use professional camera and later manipulated the footage in post to look like that [meaning like a smartphone footage]. To be honest, that would always be my choice, because it's a much easier way to manipulate and handle the footage. The smaller cameras have lower recording bandwidth[x], and that defines the look. If you're looking for something more aesthetically beautiful and filmic, you can't get that with a small frame (Elsie).

The insistence on the higher quality of the footage was equalised with his aesthetic preference for widescreen, but Roi also speculated the perception of acceptable footage quality is related to the generation of content producers.

> If I were younger, in my 20's, the quality wouldn't matter so much to me. I saw it in my students. They accept differences in footage quality much easier as they are used to mixing a lot of different resolutions, and aspect ratios. Their brains there are more used to it (Roi).

Such comments offer stimulating insight into generation gap stereotypes as Anna argues that the younger audiences still care about footage quality, but, unlike the older generation, millennials does not have a strong preference for screen orientation. As a testament to her pursuit of quality, Anna pointed out that her TVNZ team uses the Canon cinema line of cameras, C200 and C300. Moreover, whenever possible multiple crew members go on the shoot in support of the principal producer, so they together deliver better quality footage. 'The default capacity expected within the team is that everyone should be able to operate as a single person, but we are encouraged to work together because you get a better quality video' (Anna). Anna also indicated she thinks the contemporary online audience is highly aware of the image quality and has various tolerance levels concerning type and length of content.

> An audience can forget a shot that's a bit ugly if it's vitally important; for example, if somebody live streams a police shooting. In contrast, in my unit, we find that audiences aren't willing to forgive really bad shooting on major pieces if they're going to spend 10 minutes watching YouTube on a computer or even on their phone. In that case, they want a video to look good (Anna).

Even though camera technology progresses rapidly, implementing upgrades sometimes takes more time in professional systems than in consumer devices. Despite preferring larger cameras, Roi acknowledges technical improvements of a smartphone camera and points out that sometimes their characteristics outperform more expensive devices. To promote theatre plays his group makes, Roi first started producing 60 seconds vertical teasers for social media in 2014. It was also the first time he used a smartphone as a principal recording device combined with a full-frame DSLR camera for a major project.

> I bought an iPhone for the 2014 show, since my troupe needed to film a high-speed slow motion but also to transmit a wireless video feed live during the performance. It was the first time I considered a smartphone as a recording device because there was no other on the budget option for high frame rate recording, and my large DSLR wasn't capable of that (Roi).

Despite supreme image quality and difference in the size of the sensor, the majority of DSLRs couldn't record more than 60 fps in 2014, while iPhone six was able to capture up to 240 fps at 720p resolution (Allain, 2014). After the shoot, Roi spent a significant amount of time in post-production colour grading smartphone footage to match the DSLR's look, so the audience couldn't tell the difference and stay immersed in the story.

As a counterpoint, Elsie once again challenged the use of a smartphone in a professional production environment by contrasting freelance gigs where a few multitasking crew members have ample opportunity for pre-and post-production, against the fast-paced environment featuring multiple units

and dozens of people. For Elsie, the main issue with non-professional cameras is fact they're not set up for workflows with multiple departments involved.

> With consumer devices, you usually just take the card out and put it in your computer. You don't have to think about the footage conforming[x] aspect, multiple sources you might get with different technical parameters, colour grade aspect, or what happens if you are not part of the project anymore an someone need to continue your work (Elsie).

Elsie also underlined non-professional recording devices are not robust enough to assure safe operation, and because there is no official standard, one cannot guarantee the same result each time which is critical in a professional workflow that might span for weeks and months.

> It's the ability to be sure of your workflow, to be sure of the ability of the camera to sustain repeated setups build up, break down, build up again on a different location, and the years of doing that. Also important, there is no knowledge base to troubleshoot, which can often cause a lot of delays (Elsie).

I concur with Elsie that delays can be costly in a fast-paced environment with hundreds of people on set, and the lack of standardised filming equipment can be frustrating for the team. But when it comes to mobility, it is not possible to ignore the advantages of smartphones. At the same time, while smaller physical size often comes as beneficial, having a device too compact will affect the ergonomics and influence how the operator uses it. In the case of a smartphone, the lack of a dedicated function button typical on large cameras affects the type of shots the cinematographer can execute. For example, in case the operator needs to do a precise focus pull[x] that accompanies a complex camera movement, a presence of a large focus ring with a high contrast surface containing visual cues would help the operator to hit the marks, (figure 85). Also, as Roi observed, ‘because smartphones' built-in lenses are often wide-angle lenses, the operator tends to film more wide shots’ (Roi). The benefits of a smaller form factor or image deep focus can be advantageous if that is the desired style for a story. Still, if that is not the case, a lack of high-quality optical zoom and precise focus control becomes a significant disadvantage of a smartphone, (figure 86).

Figure 85: A remote follow focus with a large control surface allows the operator great precision when changing the depth of field.

Figure 86: Professional camcorders, such as Sony FS7, commonly have zoom lenses with servo controls for precise operation.

To increase the level of control, numerous third-party manufacturers developed smartphone applications and hardware that imitate specialised functions of the large camera bodies, but as Max pointed out, while improving ergonomics and providing additional custom functions, the extra gear impedes the principal advantage of a smartphone, its mobility, (figure 87).

> In documentary filmmaking, where I mainly work, it is very advantageous if you can move quickly and you can operate the camera without getting much attention; I mean that in a very ethical way. If you can operate faster, you can change locations quickly and go to places limited to a big camera. You can currently mount any type of lens on a smartphone, including the depth of field adapters and other accessories, but I'm not a big fan of that because extra gear reduces its mobility (Max).

Figure 87: Once a smartphone is equipped with a rig providing a better grip for the operator and multiple attach points for accessorised gear, the device's mobility is compromised.

The topic of mobility against image quality is a divisive one between the professionals and it was difficult to agree to a single conclusion. As Max pointed out in a humorous manner: 'there is an adequate tool for every task; you cannot do a studio shoot with a smartphone, but you also can not take a selfie with broadcast camera' (Max). In summary, the trade-off between mobility and ergonomics against footage quality again comes down to the demands of a particular project. For Claire and Elsie, participants commonly working with cinematic content, a smartphone was a poor substitute for any professional recording device. Anna, Max, and Ernabel were more flexible in accepting trade-offs associated with smartphone footage quality in exchange for immediacy and mobility. Roi's opinion was split between the two groups as he prefers widescreen and is highly appreciative of image quality, but due to the nature of his transmedia projects, he readily employs the undeniable versatility of a smartphone.

Chapter Five: Theme 4 - Pandemic Disruption and Vertical Video Engagement Potential on Mobile Devices

The majority of data and thematic analysis up to this point suggests media professionals are reluctant to operate outside the comfort zone of their well-established workflow. Still, despite a strong personal preference and preconceived notion of what is aesthetically pleasing, some are willing to explore the potential of aspect ratios alternative to widescreen. In addition, since the COVID 19 pandemic started in January 2019, media professionals worldwide also had to re-evaluate their opinion on recording devices and screening outlets they deem suitable for the production and distribution of their work. The most apparent impact of the COVID 19 pandemic was the inability to move freely and be in the company of other crew members. In a sudden shift, professionals were detached from their traditional workspace, including specialised audio/video equipment, and forced to resort to a home environment and consumer gear. However, the level of disruption varied greatly depending on the geographical location. For Ernabel, employed in the TV industry, the lockdown meant she needed to arrange filming in her living room with a smartphone as a primary recording device. In contrast, Anna, another full-time TV professional from this study, did not have to resort to a home studio since the length of government-imposed lockdowns in New Zealand was significantly shorter than ones Ernabel endured in New York where she works. Considering Ernabel and other full-time staff of many television stations across New York City spent 18 months away from studios, advanced camera systems, and professional editing suites with little advanced planning, it is easy to understand the profound changes COVID 19 restrictions brought to the professional film and television workflows. Such major disruption urged working professionals not only to use a smartphone as a primary recording device, often the only camera they had access to, but also to actively explore the benefits of mobile recording and alternatives to widescreen with greater motivation than ever before.

The lack of planning for prolonged distance work was particularly obvious regarding the choice of a recording device since even large institutions did not have enough professional cameras to distribute them to all staff members working from home. In other cases, even if a department had available cameras, on-camera talents did not always know how to use them independently without a train crew. For that reason, smartphones have become a logical choice for many media professionals without hands-on technical skills enduring prolonged lockdowns.

> Because of the pandemic, we now shoot a lot more on our phones. And I have to say, if you have the iPhone 11, or even the newer iPhones, that's better than some footage delivered by a less experienced staff using large cameras (Ernabel).

Describing her new experience as a multitasking content creator, Ernabel acknowledged the convenience of mobile technology but at the same time highlighted challenges coming from this radical shift in her production circumstances. She found filming without the trained crew demanding and time-consuming because her attention was constantly divided between the positions of on-camera talent and the camera operator. Additionally, Ernabel realised it was harder to provide diverse footage with a smartphone and without a dedicated videographer within the same working time frame.

> It was just me and my phone at home. I'm in charge of the framing and everything else. Because I was commonly a reporter with a videographer, I never really thought about framing formally. Reporters tend to be spoiled by the videographers because we often just assume they are getting the shots we need. Now when I'm doing that on my own, it's different (Ernabel).

It is reasonable to assume more dedicated crew members can achieve higher quality results in any workflow. But this comment from a seasoned TV professional highlights the inability of a single person to self-direct their own performance and provide the diversity of camera angles, heights, and focal lengths simultaneously. Correspondingly, a widescreen composition will further challenge the capacities of a single multitasking show host if a person needs to arrange a long horizontal axis with content related to the presented topic each time. Without a studio crew, Ernabel had to dress her own backdrop, and considering she was the only person on camera, a vertical frame became a more logical choice for social media show previews. As finding in the chapter four revealed, there is a minimal space for anything else in the composition once the smartphone camera was set to the vertical medium shot of a person, and without the need to articulate the entire 16:9 widescreen composition, Ernabel was able to find more unique backdrops within the same home studio space. Considering those circumstances, a shorter horizontal axis of vertical video that Roi described as self-centred and restrictive to camera movements becomes an advantage for a reporter working solo from home during a pandemic lockdowns.

Additionally, the coded intimacy of vertical composition combined with the direct address method brought similar advantages to Ernabel's show previews as it did for the single-take music video analysed in chapter four. A TV reporter and a singer both would benefit from the coded meaning of a vertical frame, even more so if the act was delivered during an uninterrupted take. As Max summarised, 'vertical for me means having a focus on talent and people. It is easy to fill the entire frame with a face or the whole body because of its orientation' (Max). Max also believes that if the audience sees a story as a vertical video, it will be easier to believe that the screened event actually happened and was not staged. He further develops his discussion on the coded meaning of frame orientation with the argument that the choice of the camera facilitates specifics of the production approach and stresses the use of a smartphone as the main camera is no different.

> Not every story is a good mobile film, but if you're working with communities, if you're working towards self-reflective filmmaking, or creating personal stories, you can do lots of things that you can't do with a broadcasting camera (Max).

A sentiment that vertical aesthetics speaks directly to the genre of content an author is creating, and that shape of the frame relates more to personal and intimate stories echoes the general audience's perception of a vertical frame described in chapter two during the literature review. As an academic and freelance documentary filmmaker, Max is well aware of formats remediation, and he acknowledges the necessity of tailoring visual content for each online distribution outlet if an author wishes to increase the reach of content. However, he also makes a clear distinction between mobile and theatre viewing experience, stressing that both are unique and there is no reason to compare them directly. Anna similarly acknowledged the potential of a vertical frame as well as the necessity to create a unique experience for the mobile audience.

> We have been filming like this [Anna makes a widescreen gesture] because that's just the way you do things on TV. Whereas if we are actually honest to ourselves about the user experience,

> probably most people are watching stuff on their phone, and it is nicer to have that vertical frame (Anna).

Despite recognizing the importance of mobile audience and the engagement potential of vertical video, in the rest of her answer, Anna returned to her previous argument underlining the limited abilities of a short horizontal axis to house multiple subjects, which in turn compromised broader integration of vertical format in her department's workflow. In contrast to Anna's and Max opinions, Claire never saw a mobile screen as a unique experience. She stayed adamant the audience can not be fully immersed in the story because the smartphone viewing experience has many distractions. 'If it's a really good narrative story or a good documentary, you want to be really focused in a movie theatre. I just don't think that that was going to happen in a coffee shop' (Claire). It is not possible to separate the subtext of Claire's strong preference for cinematic content and antagonism towards smartphone mobile screen, but it is true a watching experience on a smartphone is often coupled with multiple audio and visual distractions. In contrast to the cinema, home TV projections usually have multiple sources of visual and audio distractions, while smartphone watching experience endures even more disruptions from a screen movement, location noises, and various application notifications pop-ups. But what Claire sees as a disruption, Max identifies as a potentially unorthodox transmedia experience.

> What happens if you watch the film on location and you see something related to the story on the screen that is in front of you? It might be captivating to create these little connections in your mind while still watching a video on your smartphone (Max).

Max expands his argument by highlighting the creators' ability to do post-production on a smartphone while still present at the filming location. Such experiences are not new since high-budget productions always have a post-production crew on the filming site delivering dailies[x]. Still, such a crew commonly operates from a mobile studio and doesn't make editing decisions while present in the actual filming environment where the crew is still exposed to physical sensations such as air temperature, smell, sounds, etc. Max considers the editor's immersion in the filming environment as beneficial for judgments on the footage's aesthetic qualities and stimulating for editing sequencing. Following a similar logic, Roi embraced the ubiquity of smartphones and refashioned what was traditionally perceived as disruptive audience behaviour in the performance space into an instrument of immersion. As an example, he talks about a transmedia theatre play he wrote on the mysterious disappearance of Amelia Earhart. The whole narrative travelled between the actors on the stage, a large projecting canvas in the background, and the messages they receive on their phones. Roi's technician sent text messages and groups of vertical and widescreen video clips through an app, so the audience didn't see the crew working on the stage between the acts because they were paying attention to their phones (figure 88). 'As a result, the narrative wasn't interrupted; it kept going on the phones until the actors on the stage were ready for the next part' (Roi).

Figure 88: The disappearance of Amelia Earhart inspired Roi's transmedia play. He used a smartphone as a viewing platform in combination with the main stage canvas. Roi also combined multiple widescreen shots within a single vertical composition.

Despite stating a personal preference for widescreen on multiple occasions, Roi concluded the interview with remarks he is open to experimenting in the future with mobile technology and the expressive potential of the vertical frame. 'In 2014 or 2015, I only produced vertical content to promote plays; it wasn't creatively oriented. But now, in 2021, part of the projects are specifically meant to play on smartphones in the vertical frame' (Roi). This statement suggests the demands for vertical videos conditioned Roi to adopt a format and change his traditional attitude in exchange for a better chance to engage a target audience. Like Roi, in the conclusion of her interview, Ernabel expressed uncertainty about the broader adoption of vertical format in the Broadcast television industry, but she admitted the growing importance of vertical video as an audience engagement tool and expressed commitment to further exploring the format.

> I don't know if I could have more vertical videos in my magazine show without some extra resources. But I do want to teach about a vertical video to the generation of students who are the vertical viewers. If that's the way they see the world, that should be the look of the content they create. I think that's important if you're an educator in media (Ernabel).

Interestingly, when given a chance for her closing remarks, Anna's response was similar to Ernabel's. 'I would just add that this conversation has made me realise that I need to think more about aspect ratios of our online content, because we need to prioritise tailoring the content more for each platform' (Anna). Having multiple participants expressing their increased interest in a vertical video after our interviews makes me feel like an advocate for the format. That was not my original intention, but by provoking working professionals to look at the potential of vertical format with a fresh pair of eyes, I felt I was challenging the aesthetic barriers I once considered definitive before embarking on this research.

Chapter Five: Talking to Professionals – Conclusion

After processing the interview data, the first conclusion that came from the thematic analysis was that vertical video is still a divisive topic between media professionals. Regardless of my best efforts to stay away from direct inquiry correlated to preference, most interview data was imprinted with a subtext of participants' personal inclination towards widescreen aesthetic. Multiple statements echo the sentiment of the initial backlash against the vertical content and serve as a reminder of Canella's point, introduced in chapter two literature review, that conventions of visual aesthetics are firstly established in film and journalism schools and nurtured between colleagues in the workplace. For that reason, personal predisposition for screening aspect ratio can significantly influence creative decisions in the professional environment, and numerous statements from the acquired data support that.

Furthermore, most of the participants agreed mobile screens became a vital distribution channel for any form of moving images, but despite being aware of performance markers on selected social media platforms, Anna and Ernabel pointed out they still used vertical video occasionally and never invested too much effort in the original production because they do not perceive the format as a separate entity, but rather a supplement of the primary widescreen piece. Such attitude corresponds with the findings from chapter four, where vertical versions of music videos commonly serve as a booster to original widescreen segments. Anna and Ernabel also blamed the absence of resources in their departments as another reason for the lack of original vertical content in program they produce. The inability of a short horizontal image axis to house multiple people in a medium shot or larger number of characters necessary for closed caption, proved to be another fundamental reason why TV professionals are not using it more for their social media cutdowns.

Accepting the importance of the mobile screen due to its immense reach and vertical video as an engagement tool, Claire, Elsie, Roi and Max also agreed that if vertical content was produced from the ground up for a particular platform the results were much better than any re-purposing of original widescreen content to a slimmer frame. For that reason, a professional crew will have to adapt the arsenal of their skills to the vertical aspect ratio. That involves the entire range of above and below-the-line crew, from directors tailoring scripts to the canvas of a different shape and size, to editors re-thinking conventions of the language of moving images as they are looking for an appropriate moment to cut.

It was difficult to reach a definite conclusion on the topic of the advantages of mobility against the image quality and ergonomics in the interview data, because there are so many variables in media production that need to be assessed separately for a more definitive answer. If the mobility of the primary recording device is the highest priority, as Max often pointed out referencing documentary production, the question of lower footage quality becomes secondary for the author interested in capturing moments that can not be staged or repeated. As a counterpoint supported by Claire and Elsie, if mobility is not an imperative, and there is an available crew and budget to use a professional camera system, the use of a smartphone as a primary recording device is not justified in a professional workflow

since footage can be manipulated in post-production to resemble the look of the smartphone. In addition, being a non-standardised filming device, the use of a smartphone creates multiple production and post-production difficulties, that as a result generate additional cost that questions its use in the first place.

In many ways, content production from the home office during the past two years was the only way for many professionals to continue to work. Once faced with restrictions governments imposed as an answer to the COVID 19 pandemic, media professionals had to re-think their views on acceptable recording devices, image quality, filming location, and distribution outlets. Once separated from the comfort of a professional environment, and detached from advanced video, audio, lighting and broadcast equipment, media professionals all over the world had to quickly rethink ways to produce content without specialised gear. Once they were forced to use all of their skills in combination with smartphones, often the only video recording technology at hand, they were also compelled to re-visit their opinion toward vertical video aesthetics.

Lastly, reflecting on my personal experience as a film and TV professional, I have to admit I shared a similar attitude toward novel formats. While being employed full time in the TV industry and as a freelance director of photography, the pace of work and daily routine often left little or no time to explore novel formats. Moreover, reflecting on professional training is not an easy process, as many technical and aesthetic decisions are made subconsciously after doing them for years in the uniform production environment. Only after leaving the industry and with prolonged exposure to other kinds of content creators, particularly from a younger generation, I was able to let go of certain rigid rules like the rule of thirds, dismiss them as axioms and see vertical video as an expressive format with fresh eyes. Challenging key fundamental principles is always hard, but it comes down to the willingness to explore new methods, and sufficient knowledge to select the best filming tools and aspect ratio to tell a particular story.

Chapter Six: Creating Artefacts – Introduction

Through the creation of four music videos, two vertical and two widescreen thesis artefacts #1, #2, #3 and #4, in this chapter, I empirically tested observations from the textual analysis conducted in chapter four and comments from the interview participants thematically analysed in chapter five. Throughout the filming process, I scrutinised benefits of a smartphone mobility in creating vertical content against its other trade-offs such as inferior footage quality, prevalence of deep focus and ergonomics when compared to professional camera systems. Thinking of a mobile phone as a primary device for vertical video consumption in the present context, I will also scrutinise how the vertical canvas and smaller screen influence the technical and aesthetic decisions of a content creator. I chose to create music videos for solo musician, Mr Russel Henderson (artefact #1) and a 10-piece reggae band, Ras Judah & Culture Embassy (artefact #3) in order to explore the contrasts and challenges of shooting a single subject versus multiple subjects within the vertical frame. As I was aiming to critically reflect on the benefits, and limitations of music video production within the creative paradigm of vertical filming, I also filmed alternative widescreen take of solo (artefact #2) and band performance (artefact #4) as a reference.

Chapter Six: Creating Artefacts – Filming With a Smartphone

With a career starting in the 2000s, I was able to witness the remediation of analogue toward digital technology, and the profound changes that this shift brought into the workflow of film and video production. Today, twenty years later, there is a broad range of consumer and professional devices able to record a high-resolution video file, and even if a device does not have any pre-set option for filming a vertical video in the system, essentially any camera mounted on support heavy enough to counter its weight can be flipped for 90 degrees to record vertical footage. Furthermore, digitally recorded material can be manipulated in a range of free and specialised non-linear editing programs to fit the frame of any shape. Different benefits and restrictions are coming from the use of a professional camera system that needs multiple crew to operate at the optimum speed, against new generation smartphones such as iPhone 12mini that can fit in the pocket. Affordability of the system, as well as mobility and device form factor, can affect the filming workflow in numerous ways, but ultimately there is no machine that can compete with the convenience of a smartphone as an integrated recording, post-production, playback and distribution device.

If I take the convenience of a form factor aside there are still two key distinct characteristics separating smartphone footage from the material recorded in medium size or a large camera body, one technical, a combination of footage colour sampling[x] rate and bit-rate[x], and the other aesthetic, the depth of focus. Reflecting on the debate on smartphone image quality from chapter five, larger camera bodies commonly deliver higher image quality, but the improvements in footage quality always come with the increased device cost as well as necessary storage space. Primary benefits of larger file sizes come from a higher chroma subsampling and bit-rate that improves the accuracy of colour reproduction in the video signal, while simultaneously providing more creative possibilities for colour grading due to a higher dynamic range of footage (ITU-R, 2015). Also important, the size difference between the sensor surface of a smartphone and a full-frame camera can be between 15 to 50 times depending on the manufacturer, and physical size of a sensor is another key technical element determining the quality of a recording (figure 89) (Dempsey, 2013).

Figure 89: The comparison of the sensor size difference between a smartphone, point and shoot camera, small DSLR, and a full-frame DSLR.

In contrast to large cameras, advanced algorithms and data compressions are applied in smartphone operating systems to improve the image quality while preserving a smaller file size, but when it comes to the prevalence of deep focus in the recorded footage, that is not a matter of choice. A combination of a small sensor and tiny lens is responsible for extremely deep focus in the default settings of any smartphone footage, (figure 90).

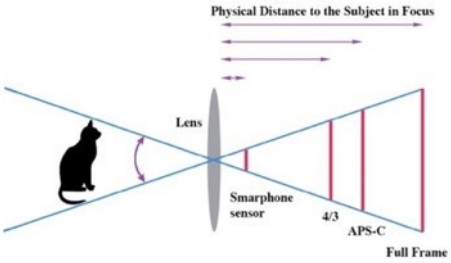

Figure 90: Diagram illustrating how the size of the sensor influences the depth of focus in the footage.

As a result of this deep focus, the spectator's gaze is 'wandering around' in search of movement, eyes, bright colours or high lighting levels, and if the deep focus is combined with complex composition, this results in the shot needing more attentiveness to process (Smith, 2011). In order to change the aesthetic look of the footage, a smartphone algorithm can introduce a filter that blurs a section of the image on user prompt, simulating a shallow depth of field, but depending on the shape of a subject, the filter is not always accurate its use becomes questionable in a professional workflow.

Complex multilayer compositions along the Z-axis are nothing new, and well-known examples such as Orson Welles *Citizen Kane* (1941) are part of cinema history (figure 91). To achieve extremely deep focus in the 1940s, cinematographer Gregg Toland had to build parts of the set oversized or smaller than the actual proportion, and install a fake cloth ceiling at an angle to achieve the effect (a technique called forced perspective). Lastly, Toland needed an abundance of light on set to provide enough exposure for the lens stopped down at a large aperture number (Bordwell, 1985, p.319). Because smartphones are aimed at a consumer market, it is safe to say handset manufacturers were less interested in the cinematic expressive potential of shallow depth of field, in exchange for the possibility that an everyday user easily records in-focus footage from foreground to background. In conclusion, while being easier to operate and delivering technically correct footage, the smartphone as a primary recording device offers fewer aesthetic possibilities than a DSLR, SLR or large camera body in respect to the depth of field.

Figure 91: Composition with an extremely deep focus from Orson Welles Citizen Kane (1940).

Another critical element in a professional environment is the speed of file transfer and the energy necessary for the process of filming. With no option to quickly swap recording media, it is not possible to transfer video files quickly from a smartphone to a computer or external hard drive. A similar issue applies to batteries, since even the speed of fast chargers is unacceptably slow for professional workflow where a fresh camera battery can be changed within a few seconds. During her interview, Elsie gave an example of other problems coming from the use of unstandardised workflow and a smartphone on a professional set. While filming Wellington Paranormal, the production used a range of cameras and smartphones. Each camera was selected for its ability to deliver a particular look. ARRI Alexa was a principal recording device supplying a cinematic looking footage, small GoPro's represented a police officer's body cams, and a smartphone filming vertically provided an average users point of view. As Elsie explained, the use of various cameras was partly justified but not necessary as any look can be made in post-production from high quality recording. As a trade-off, the inability of Go Pros and smartphones to generate standardised timecode and metadata in the footage created a series of problems in post-production: 'The main problem we have with not using large format or any other professional standard is the ability to push it seamlessly through pipelines (in this case the interviewee refers to post-production pipeline), there's always problems with timecode and tape names' (Elsie D). Inserting timecodes and manual synchronising of the footage without standardised tape names in non-linear editing software is a tedious process that increases the possibility of mistakes and generates additional post-production costs. As Elsie concluded, it comes down to the directors' preference and available budget, if production will actually use an amateur device or create the look in post-production from the footage recorded with a professional camera body.

Since I filmed thesis items #5 and #6 with a larger camera body and the help of a crew (details are available in chapter seven), I wanted to juxtapose that production experience with a set featuring recording device that is accessible to anyone, can be controlled without any specialised training, and operated by a single person, that is a smartphone. Here is the summary of all artefacts produced as a part of this research with hyperlinks to the files, (figure 92).

Item #	Name	Description	Recording Device
1	Saxes in 722 - Vertical	A single take vertical music video featuring a solo talent.	iPhone 12 Mini
2	Saxes in 722 - Wide	A single take vertical music video featuring a solo talent.	iPhone 12 Mini
3	Mumandi - Vertical	A vertical single take music video of a multi-piece band.	iPhone 12 Mini
4	Mumandi - Wide	A vertical single take music video of a multi-piece band.	iPhone 12 Mini
5	Loss Vertical - Film	A short vertical fiction film, edited in 1:2 aspect ratio from the original 16:9 widescreen material.	Canon C 300 with Sony CP2 lenses
6	Loss Vertical - Trailer	A 30 seconds vertical trailer, in 1:2 aspect ratio.	Edited from item #5

Figure 92: Summary of thesis artefacts.

The smartphone that I used for filming of thesis artefacts #1, #2, #3, and #4, was an iPhone 12 mini, featuring an A14 Bionic processor able to handle High Dynamic Range[x] (HDR) video signal on a sensor level, and main wide-angle lens with minimum aperture of f/1.6. As a comparison, any photography or cinema lenses in a professional full frame line up with declared speeds at f2.8 or less, are considered very fast. Those specifications are sufficient to process 10bit video signal at 4K resolution, and to compress the video file with HEVC (High Efficiency Video Coding), a successor to the widely adopted H264 developed for online video (Apple, 2020).

So why are technical specifications of a primary recording device important? Firstly, the faster lens can direct more light toward the image sensor, and second, a device with more processing power can interpret received signal better and subsequently pack the information into a smaller video file container for easier online distribution. This comparison is essential as the majority of high-quality DSLR and mirrorless camera bodies, often costing 3-5 times more than a smartphone (price with included fast lenses could be 4-7 times higher), still don't have enough processing power to record HDR video. Top of the line cameras from manufacturers such as ARRI, RED, Black Magic and Sony are able to acquire HDR signal, but the cost of these systems and the size of the files they deliver requires significant investment in the media cards and post-production equipment, so it is available only in high budget productions. This relationship between advanced HDR algorithm available in a smartphone but not in many other more expensive cameras is analogue to high-speed recording possibilities iPhone six had in 2014, and Roi mentioned in chapter five, beating technical capabilities of any DSLR available at the time.

Having the benefits of mobility, I still wanted to make sure any vertical and reference widescreen footage I produced wouldn't exhibit external disruptions that might influence audience perception, or mine during the analysis. Bumps and camera shake, alongside any other unmotivated secondary motion, draw awareness to itself and upsets the attention synchrony of the audience.
For that reason, I opted to use a DJI OM 4. DJI OM 4 is a foldable three-axis stabiliser designed to work with a range of smartphones, and equally important for this research, OM 4 has a built-in function that allows the operator to quickly switch between the horizontal and vertical filming, (figure 93). The ability to change filming orientation fast, meant the adaptation of the stabiliser would not cause any delays between the takes that might influence the performances. Lastly, servo motors built into the stabiliser allowed me to execute corrective camera pans and tilts at designated speed, and repeat the movements with greater precision than would be possible handheld.

Figure 93: DJI OM 4, smartphone stabiliser.

In a digital media production workflow, options provided by software used for filming are equally critical as the capabilities of the phone camera and support hardware. Proprietary applications in the phone operating system often have only basic functions, and an interface designed

for simplicity of use. Because default software does not commonly offer manual control of exposure and other filming parameters, in order to use the full potential of available hardware a smartphone user has to install specialised software. Such specialised software is not free but also is not too expensive, ranging from a few dollars up to a one hundred for higher end options. With paid software such as Filmic Pro or Carousel Camera installed, the user gets manual control of aperture, shutter speed and ISO, an option to use the full dynamic range of the smartphone's sensor, control over bit-rate and footage resolution, as well as features such as audio meter[x], histogram[x], and colour scopes[x] necessary for accurate monitoring of the recording.

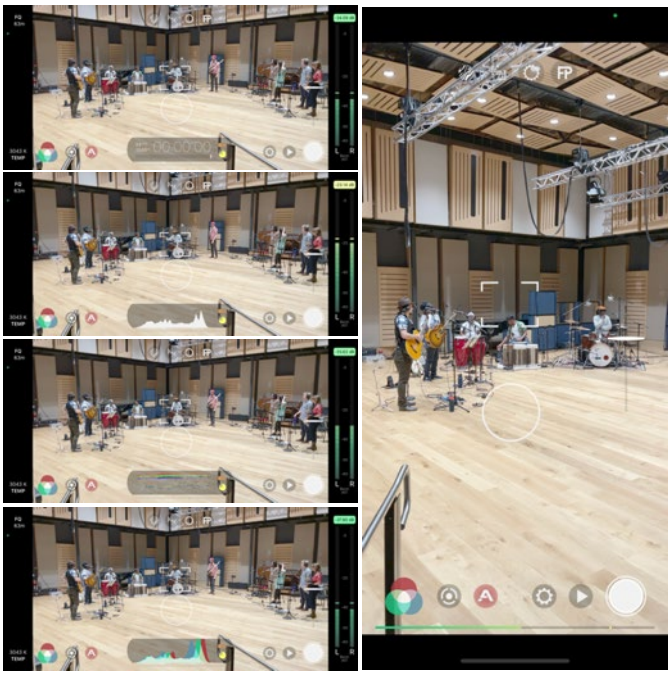

Figure 94: In widescreen orientation, Filmic Pro offered a few different versions of the footage histogram. Once turned to a vertical orientation, the application GUI was significantly changed, and with no visible histogram.

I also noticed the interface of the Filmic Pro application looked different depending on the orientation (figure 94). Manovich in *The Language of New Media* (2001) emphasised the influence of GUI on the user by discussing graphical user interface as a creative tool, but also as potentially limiting factor in creativity since the individual might explore only options presented in the platform (Manovich, 2001, p.75). The primary reason I opted for using Filmic Pro over DJI Mimo application were its advanced options and the more professional graphical user interface (GUI), but in a vertical orientation, Filmic Pro did not include a histogram nor audio meter. A constant monitoring of audio and luminance levels are mandatory in any professional filming workflow, and lack of colour scopes opened more space for potential mistakes. At the same time, the minimised icons and consequent lack of shifting parameters visible in the scopes made the screen less populated, and monitoring of principal action easier. I couldn't find the definitive answer on companies website as to why the GUI of Filmic Pro looked so different between vertical and horizontal orientation, but I assumed vertical alignment was attuned for video blogging, where constant movements of histogram graph might create a visual distraction for a person sharing the roles of an operator and on-screen talent.

Lastly, on a physically small device such as a smartphone, where touch screen space is contested between image preview and camera controls, the lack of designated physical buttons will hamper the operator's ability to execute accurate focus pull. Furthermore, touching the screen necessary to change the focal point in the footage would compromise the camera movement with unwanted bumps, (figure 95).

Figure 95: Third-party applications such as Filmic Pro offer a programable focus dial, often inaccurate enough for complicated camera movements where focal points are changed multiple times.

Because the specifics of filming location are vital for the overall look of visuals in any video production, the last pre-production step was a location scout of the studios. For vertical filming, the relationship between the width and height of the space is even more important because it dictates the maximum camera distance from the subject before the ceiling gets into shot. This is not a novel issue, since over the course of film history, changes in image aspect ratio meant the studio spaces can be used differently, depending on the shooting format. When cinematographers Arthur Edeson (*The Big Trail*, 1930) and Robert Planck (*The Bat*

Whispers, 1930) planned composition of the scenes they filmed with a 70mm camera in the then newly established 2.10:1 cinematic aspect ratio, the size of the format and shape of the canvas influenced several aesthetic decisions.

> Both cinematographers frame their scenes with lower heights, as certain widescreen traits become apparent—specifically, challenge to traditional Academy ratio blocking strategies, camera movement, and camera angles. The lower camera height led to fewer editing cuts and prescribed other adaptations of traditional filmmaking trends, such as lower overall set heights and lighting strategies that are more lateral as opposed to deep or vertical (Cossar, 2011, p.65).

Figure 96: Studio at Massey COCA.

For vertical filming, the height of the ceiling proved to be particularly important, as a vertical frame gives away the source of the light on the overhead grid sooner than widescreen. This was a problem I faced while filming both performances but exacerbated with a band, and the issue associated with vertical video mentioned by Anna, Claire and Ernabel in chapter five. For example, since a vertical shot wasn't able to encompass all musicians at once, a descriptive pan was necessary to introduce the band members at the beginning. In an effort to minimise redundant camera pan I moved as far as possible from the arc of musicians, back at the studio door, but that position brought an even larger aesthetic compromise as an unnecessarily wide shot encompassed doorway safety rails, as you could see in figure 96.

Figure 97: Studio at Victoria University School of Music.

In a similarly spacious studio rented to film a single jazz player, I had to be mindful not to leave a large area of unarticulated space in the composition, as that would draw too much audience attention in the sequence with fewer visual accents than the band was able to produce. Having a single performer on any large stage would require additional planning on how to articulate the empty space in any widescreen composition as well, but more prominent presence of floor and ceiling on account of commonly broader lateral space in my vertical shots required different pre-production work. The pre-production time spent in space preparation for the filming of artefact #1 and #2 was also a moment to reflect on Roi's comments from chapter five, that filming of vertical content in spaces traditionally attuned to widescreen compositions requires an additional creative effort. On that account, while preparing the studio, I spent significantly more time masking horizontal distractions such as cables and other unused studio props, but I also had to consider how lighting design for a musician position works in combination with visible overhead lighting sources and floor reflections that I could not avoid in wide vertical shot. After several hours of work, the relationship between positive and negative volumes, together with interplay of light and shadow, produced a stage-like atmosphere that worked well for a jazz performance. Furthermore, a single performer positioned in the strongest pool of light had enough separation from the background to become a principal point of interest, and the character of Mr Henderson's music worked well with significant 'breathing space' surrounding him as you can see in figure 97. Finally, to balance visible filming space below and above performer, I set up approximately 120 audience chairs to create 'visual corridors'. Despite not having an actual audience, the presence of vertical and horizontal graphic vectors on the floor balanced the visual presence of studio architectural elements above, and improved the perception of 'off-stage space' that worked well for the character of music in artefacts #1 and #2.

Once filming gear and the space were ready, I finalised the planning of the camera path. Instead of looking for variety in the coverage, heights and angles, I filmed the artefacts with the idea of having the entire performance in a single, uninterrupted camera take at eye level. That method resembled the aesthetics of the key online samples I analysed in chapter four. A single take aesthetics also allowed musical accents in performances to become cues for camera movements so the rhythm created by editing wouldn't take attention away from the shape of the shot. Keeping in mind a breakdown of the footage characteristics found in the key samples of the vertical music video (presented in chapter four, figures 36 and 38, page 39), I aimed to spend roughly 60 percent of the performance running time in

medium, 30 percent in close-up and ten in wide composition. Lastly, because a single uninterrupted take is commonly perceived as being a more factual account of the event by the audience, it was an aesthetic worth following. After I started filming, I needed one or two takes to identify key musical accents in each track and adapt the camera movements to their rhythm. In the case of a single performer, all of the cues came from the musical accents produced by the main instrument, baritone in the first video and alto saxophone in the second.

Chapter Six: Filming of Artefacts #1 and #2 / Solo Performance

The first set of screenshots offers an illustration of the principal differences between cropping and the use of alternative aspect ratios during the filming, (figure 98). As concluded in chapter four, a simple crop is commonly a poor substitute for the original mise-en-scene, since a newly created composition often carries a different meaning.

Figure 98: Vertical 9:16 crop of the original 16:9 footage.

Figure 99: Illustration of the camera path, from the left corner, across the stage area into the right corner and back along the same route. Start and end positions are visible in the reverse shot.

In contrast, thinking ground up about the relationship between action in front of the camera and attributes of a frame opens more creative possibilities to explore a specific characteristics of the filming aspect ratio. Considering mood of the performance, lighting design, as well as size and shape of the studio, the camera path started in left corner of the space leading towards the musician into a half-circle around him, and ending down the aisle at the right corner. I paused for several seconds in every key position, waiting for the switching accent in music to motivate the camera movement. From the right corner, I followed a similar arc back (figure 99). Slow camera movements in combination with subtle corrective pans and tilts corresponded with those observed in key vertical samples analysed in chapter four. Following a simple camera path also meant I could repeat the movement with greater precision and have a more accurate comparison between widescreen and vertical take. After a few tests, I started with filming, making sure I changed the screen orientation each time Russel began a new take.

The artefact #1 commences from position 1, with the camera hidden behind a pillar for the starting reveal of the performer. The moment of the reveal appears to be better motivated and timed in the vertical arrangement due to rapid changes in visible elements. Still, on a wider canvas, I could incorporate more reflections in the stage area, making both compositions compelling, (figure 100).

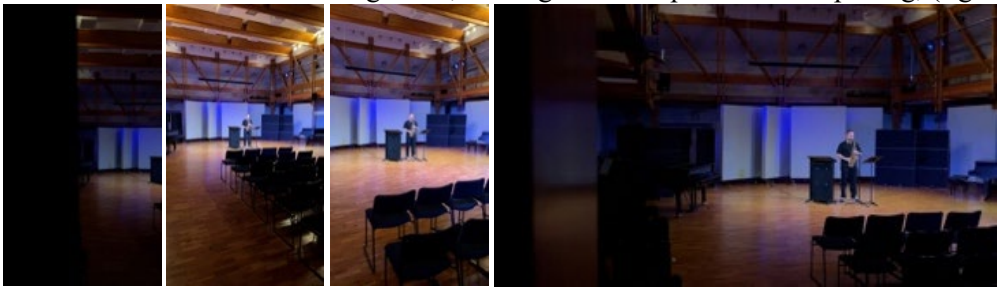

Figure 100: A comparison of a vertical against a widescreen shot from position 1. By encompassing a large section of floor and ceiling, the vertical frame rendered a feeling of a larger space.

As I slowly approached the subject, the rows of chairs stayed visible for a longer period of time in the vertical shot, as well as the structural elements of studio space, including lighting sources between the beams. If this performance had an audience, a vertical shot would work well for the start as the musician would share the same space with an audience for a longer time. Also important, because this particular space featured intricate architecture and lighting design, the interplay of colour, shadows and light produced a pleasing overhead canvas for the performance. I was aware not every filming space has an aesthetically pleasing ceiling and such a large volume, and for that reason I decided to further explore the potential of this 'taller' studio space in my compositions. If I was filming in a space where the look of the ceiling wasn't as pleasing, or it would cause an audience distraction with its look, I would be compelled to stay closer to the subject and avoid very wide compositions.

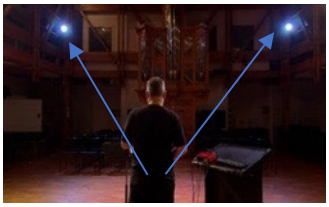

Figure 101: Balanced composition with strong hints of graphical vectors extending into off-screen space.

While the camera travelled behind the performer with the intention to show a reversed angle that includes the musician and the audience, I paused waiting for the next accent (figure 101). This was also a good spot to deliberate the difference between two aspect ratios. The aforementioned aesthetically pleasing ceiling was omitted from the widescreen shot, but the audience section had equal prominence unlike in the vertical one dominated by the performer. Furthermore, during the seconds where both lights stayed visible in the widescreen, the triangle between them and the stage area produced a balanced V-shaped composition that projected 'upward' frame intention, crafting a positive subtext for the music.

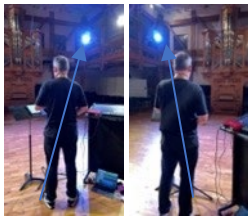

Figure 102: Unbalanced composition induced by the diagonal graphical vector between musicians body and lighting source.

In contrast, accents created by stage lights produced a less engaging diagonal composition in a vertical frame since I wasn't able to incorporate both overhead units at the same time. Moreover, with less physical space on the left and right-hand sides, the vertical frame was dominated by the musician's back, rendering a less engaging composition without much audience or potent graphic vectors, (figure 102). Moving from the centre position 2 towards the end position 3 at the far right corner of the studio, I resumed framing similar to the starting one and waited for the bridge in music to start the next movement, (figure 103). While tracking back to the middle of the stage, instead of going behind, I stayed in front of the player, holding an MCU shot to emphasise the instrument. Because I couldn't encompass the entire instrument in a 16:9 composition, I had to introduce a slow camera tilt, adjusting the frame from the player's bust to a shot encompassing only his hands at the lower part of the saxophone, (figure 104). In contrast, the vertical 9:16 frame incorporated the player and his entire instrument so there was no need for the camera movements, (figure 105).

Figure 103: A vertical view from right corner of the studio, position 3.

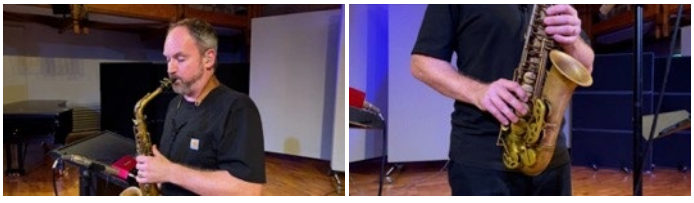

Figure 104: A widescreen shot from a similar distance produced a composition that cropped the instrument, prompting the camera to tilt to show the entire action on screen.

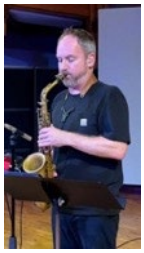

Figure 105: A vertical shot one meter away from the performer produced a well-balanced shot without requiring camera movement.

During the filming of one vertical pass, I recalled a statement from Claire she made referencing the production of vertical footage for Quibi in chapter five: 'if I were looking at you, I'm not looking at your feet, I'm looking at your face, but Quibi vertical shot could encompass all of your body. It's like

boring information' (Claire). If I relate that comment to compositions visible in figures 104 and 105, I agree the musicians' legs are less relevant for the action, but vertical medium shot accentuated the contact between performer and the instrument more than widescreen. Applying a similar logic to the rest of artefact #1 and #2 footage, if I had to choose between an empty stage area and a sense of intimacy produced by the tight vertical frame of the player and his instrument, I would always opt for a composition that evokes more emotion, which in this case is a vertical frame. After multiple reviews of the all takes, I also concluded that the look of the space played an equally important role in my preference of screen orientation as the number of subjects and type of the performance. That is evident as the widescreen composition in this setup produced a less engaging frame by incorporating other items present in the studio, but not directly related to the performer, while vertical frame encompassed visually rich ceiling that brought positive subtext for the music.

Chapter Six: Filming of Artefacts #3 and #4 / Multi-Piece Band

Figure 106: An example of the difference between the amount of headroom and floor space included in a widescreen and vertical composition, in a scenario when a camera need to capture a large group of performers spread across the horizontal plane.

The second filming location for artefacts #3 and #4 was the flagship studio of the College of Creative Arts (COCA) on the Wellington Massey University campus. The dimensions of Massey studio were comparable to the space used for the production of artefacts #1 and #2; however, the visual aesthetics of the space were quite different, (figure 106).

The ceiling structures provided an equally engaging graphic vectors, but high-key lighting and cream tonal pallet rendered a very different visual aesthetic. Conveniently, the upbeat character of Culture Embassy's music fit well with the abundance of light where every facial expression and communication between musicians stayed visible. Knowing Culture Embassy's performance would have more visual accents than a solo performer, I spent significantly less time hiding cables, stands, chairs, and other studio props from the shot, relying on the busy mise-en-scene and camera movements to occupy the spectator's attention with ease.

Figure 107: The size and shape of the instruments, as well as the drummer's position, were much better suited for a widescreen shot than vertical.

Figure 108: Filming elongated instruments, such as trombones, in combination with wide movement along X-axis necessary to play them, was challenging in a vertical frame.

One recurrent issue during the filming the band vertical takes was my inability to quickly find a balanced frame while filming multiple performers, particular groups of sitting drummers, as a slimmer aspect ratio did not work well for action happening in an approximately square area. In contrast, a longer horizontal axis produced an aesthetically pleasing and balanced composition without the need to cut subjects along unorthodox lines, (figure 107). The issues I encountered during the filming of artefacts #3 and #4 were equivalent to problems Netflix trailer editors encountered when re-cutting dialogue scenes mentioned in chapter four, as well as the principal reasoning of Ernabel and Anna cited in chapter five for not using vertical composition more often in TV production. Another major problems coming from the short x-axis were limited 'safe zones' for the instruments moving quickly along horizontal plane of the composition. Having an additional leading space in front of the subject is common practice in cinematography, especially when the camera operator films an action, not knowing all movements in

advance. The issue of safe zones was exemplified while filming musicians playing trombone, trumpet and guitar due to the fact that their bodies were in vertical orientation, but their instrument moved in horizontal alignment. Lastly, since short horizontal axis couldn't provide enough space to compose a shot with well-defined direction intention, I opted for wider vertical shot encompassing groups of musicians, as that arrangement provided more balanced composition despite additional floor and above head space, (figure 108). Knowing that articulation of the Z-axis can be a solution for shorter horizontal axis, I attempted to create a multi-layered vertical composition by arranging several musicians along the depth of the shot. Such angle only occasionally produced the desired effect as you could see while watching artefact #3, but more often, the composition looked too busy as deep focus with elements moving in foreground, middle, and background made a separation of visual accents impossible, (figure 109).

Seeking to provide a stronger separation in the composition through the depth of focus, I occasionally moved physically very close to some musicians and found a smartphone camera was able to produce some background blur from a shorter distance, but not enough to highlight a single subject. For example, because of the deep focus I couldn't single the profile shot of the guitar player during his solo (figure 110), or get a CU shot of his fingers on the strings due to a lack of optical zoom on a smartphone. Additionally, such proximity made some performers, like a trombone player, uncomfortable as I was at risk of obstructing broad movements of their instruments'.

Figure 109: Deep focus on a smartphone limited my ability to single out elements in the composition.

Figure 110: The lack of optical zoom on a smartphone was another important feature limiting my ability to provide a variety of shots from a single position.

Recalling those moments in post-production, I deemed the inability to precisely control the depth of field as a major limiting factor to the author's creativity when filming with a smartphone, since such limitation considerably influenced the footage aesthetics. The prevalence of deep focus in all shots meant I wasn't able to single out musicians by controlling the focal plane of the image. For the same reason, I needed to move more often in order to change the angles while trying to avoid background elements unrelated to part of mise-en-scene I wanted to emphasize at that point in the performance. What vertical composition also lacked during the filming of the multipiece band was the ability to frame a wide shot encompass gestures and interaction between groups of musicians. As one of the key visual accents, the performers' exchange of reactions are hallmarks of live sessions, and most of them were lost when musicians were singled out in the composition, making widescreen preferred choice for this kind of setup.

Before proceeding to a conclusion, I would highlight another creative approach that, in many ways, harvested the full potential of dissimilar aspect ratios in the filming of music performances. Parallel with data analysis and thesis write up, I continued to review new examples of vertical video implementation in a professional production context, and in June 2020 I found a captivating music video *Nick Cave and Warren Ellis - 'Push the Sky Away' | Live at Sydney Opera House* published on YouTube. What made this text thought-provoking was the use of mixed vertical, horizontal and 1:1 square aspect ratios within a single 16:9 widescreen frame. *Push the Sky Away* featured a lead singer, conductor, background vocals, a large orchestra and a choir, so the performance accent was continuously shifting, a similar challenge I encountered while filming Culture Embassy. To record all performance accents with a single camera, the operator can focus on the different subjects during each take, but the abundance of visual accents still poses a challenge for editing. In order to show multiple key moments to the audience simultaneously, the editor Josh Milch combined various camera angles where the position and size of the split screens window suggested which one should dominate the

composition, (figure 111). While a split screen appears like a good idea to cover multiple angles of a busy scene, a few screens active at the same time can challenge the audience's perception. If overlapping planes proved to be too dynamic due to presence of multiple actions, the editor applied the fade-out effect to soften the transition between newly active and fading split-screen.

Figure 111: Nick Cave and Warren Ellis - Push the Sky Away | Live at Sydney Opera House. A selection of screen shots from the music video released in Dec. 2019, featuring a combination of aspect ratios

Figure 112: A combination of square and vertical frames in 16:9 widescreen.

The split-screen solution is not new, and great examples can be found throughout the film history, in *Napoléon* (1927) and *The Boston Strangler* (1968), to name a few. Still, what makes the use of a split screen so appealing in music videos, is the fact that the audience can see more angles of the performances and talent at the same time while the editor selects the most appropriate widescreen, square and vertical composition depending on the subject they feature. For example, a square frame and vertical aspect ratio are an excellent combination to cover the moment in the performance when Nick Cave conduct the choir while other singer takeover as a leading vocal, (figure 112). Another sample comprises three vertical compositions showing a conductor, lead singer smiling, and detail of piano keys at the same time, (figure 113). The only other way to encompass everyone involved would be a wide shot of the performance space, but in that scenario, the camera distance would hide the emotions visible in closer composition. Lastly, a split-screen allowed the audience to see multiple climaxes of live performances, as well as reaction as artists exchange ques, smiles and praises.

Figure 113: A combination of three vertical frames in 16:9 widescreen.

Concerning the difference in the size of the space and number of performers, the Culture Embassy also features one leading vocal, a few background vocals, multiple instrument groups, and equally important accents that are happening at the same time in various locations, so they cannot be included all in a single vertical or widescreen video. For that reason, I would arguably produce a more engaging music video of artefacts #3 and #4 if I used a similar editing method described in *Push the Sky Away*. The filming process would stay the same, but I am positive a mix of aspect ratios in a split-screen would render a better result than any widescreen or vertical single take I was able to make, (figure 114). In summary, the use of mixed aspect ratios in *Push the Sky Away* music

video is a testament to how quickly new visual aesthetics are evolving in the online environment, and remediation of formats can lead to new creative solutions to the limitations of the vertical frame.

Figure 114: This figure shows a hybrid between artefacts #3 and #4, as I combined the best accents from vertical and widescreen takes following the split-screen aesthetics of Push the Sky Away.

Chapter Six: Creating Artefacts – Conclusion

Despite the apparent benefits of mobility, I concluded some technical limitations make smartphones less ideal for professional video production, particularly the prevalence of deep focus, the lack of precise focus controls and lenses featuring high-quality optical zoom. Due to a lack of high-quality optical zoom and the means to isolate a single subject along a shallow focal plane, I had to move significantly more during filming to provide a variety of compositions. While recording artefacts #1, #2, #3 and #4 I was able to choose a camera path and move freely in both studio spaces, but frequent movements of the camera operator were not often possible nor welcomed on every set, since they create visual distractions for the performers. Besides, due to the lack of dedicated focus control, I couldn't precisely execute any focus pulls nor form subtext by switching focal planes within a single composition. The lack of precise focus-pull controls was brought up by Roi during his interview in chapter five, next to options to expand smartphone technical capabilities with various accessorised gear mentioned by Max. But then again, they both agreed the additional load would hamper mobility, raise the cost of the smartphone as a filming system, and question the rationale for its use in the first place instead of a larger camera that already features the desired advanced options. On the positive side, comparing the production of thesis artefacts with my previous professional experience, I noticed the difference in form factor between a smartphone and large cameras influenced mobility, but also perception of me as an operator. Being the less obtrusive piece of recording hardware, the presence of a smartphone made subjects more comfortable during recording, thus more natural in their performance.

Informed by analysis in chapter four, I was well aware that 9:16 aspect ratio coupled with simplified mise-en-scene, and closer shots I was able to get with a solo musician should produce a well-defined composition with visible emotions. In contrast, filming a group of musicians in a vertical frame proved to be a challenging task. That is, a single performer filmed in an M shot and positioned centrally in a vertical frame creates an exceptionally balanced composition due to an even distribution of edge magnetism. As seen in the key examples from YouTube VEVO, all vertical single take videos had one musician as a principal point of interest. Once challenged with multiple visual accents happening at the same time, the lack of available lateral space compelled me to correct the position frequently, and in doing so, I introduced camera movements not motivated by the action in front of me but by my intention to create a balanced composition. Furthermore, a short horizontal axis populated by multiple subjects of interest, left a much smaller margin for errors caused by corrective pans than I would have in the

16:9 or any other widescreen aspect ratio. If I tried a different approach, raising the camera angle and moving further away from the group of subjects to get a full body shot, I was introducing an unnecessary amount of positive volume in visible foreground floor space. Similar issues occurred if I tried to compensate for the lack of lateral space by moving back and tilting the shot up, introducing a large ceiling section that became a dominant volume in the composition.

Reflecting on my search and the choice of filming locations, I concluded that making a single-take music video requires more pre-production time, mainly because most specialised filming locations, such as studios, cater to widescreen mise-en-scene in terms of their shape and size, often disregarding the look of the ceiling and floors. This issue was also one of the themes Roi emphasised in chapter five. Bearing in mind a wide vertical composition of a solo performer encompasses dozens of meters of ceiling and floor space, the number of aesthetically acceptable filming locations drops rapidly, and the issue of visible headroom or unarticulated foreground is only exemplified when filming a group of subjects. To summarise, multiple framing issues encountered during the filming of group performance were significantly less prominent while filming a single musician, where vertical orientation often produced a balanced composition. A tighter and less ambiguous frame revealed more emotions on the artist's face and brought up a subtext of private creative space between the solo musician and his instrument. Conversely, once challenged with multiple accents, a vertical composition felt inadequate for numerous subjects and fast switching accents, feeding into the already established narrative of an author's preference to have a simplified mise-en-scene and fewer subjects when filming in 9:16 aspect ratio. Lastly, if a music video contains numerous performers and there is no opportunity for multiple takes, like in the case of live performances and improvisation, I would argue the editing style seen in *Push the Sky Away* is an excellent solution that combines the hallmark features of vertical aspect ratio with more popular widescreen in the same sequence.

Chapter Seven: Re-Purposing 16:9 Footage to 1:2 Aspect Ratio – Introduction

Re-using original material for an alternative project is always appealing in video and film production at any scale. Firstly, because it saves available budget, and secondly, like in the case of Netflix vertical previews, it serves as an affordable engagement tool leading the audience to the principal piece. Having one of the largest catalogues of film and TV shows in the domain of global OTT content providers, the Netflix browser is an excellent source of texts relevant to this research since elongated 1:2 aspect ratio previews are all made from the material originally filmed in a widescreen format. However, the footage re-purposing process from widescreen to vertical is far more complex than creating a diagram of the visible cross-section between the aspect ratios, (figure 115), since there is no agreement that principal story-driving elements are originally envisioned within those margins. Through a process of re-editing a widescreen short fiction film *Loss* (2019) into a vertical frame and creating an associated trailer resembling the style of Netflix previews, I investigated how conventions of traditional widescreen shot scale (V/vista, L/long, W/wide, M/medium, MCU/medium close-up, CU/close-up, ECU/extreme close-up) operate in a tall frame. Along the way, I also elaborate on the reasoning behind key aesthetic and technical decisions made during the editing process, and examine how changes in composition and aspect ratio affect transformed diegetic space.

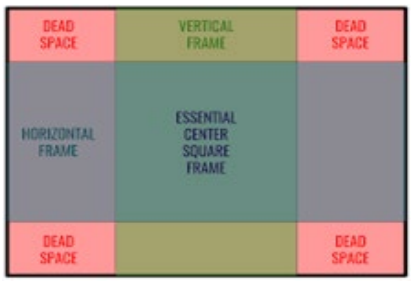

Figure 115: Thinking of a cross-section between 16:9 and 9:16 aspect ratios is a good starting point for the cropping process, but an editor needs to look beyond that area to ensure the principal action stays visible.

In the early days of cinema, when distinctive shot scales started to emerge in the language of moving images, most films were recorded on four perforations 35mm celluloid strip in a 4:3 aspect ratio. As the industry evolved and adopted other image aspect ratios and film stock sizes, the demarcation of shot scales mainly stayed the same despite the fact that the horizontal axis of the anamorphic 2.39:1 screen (the widest aspect ratio common in modern theatres) has a horizontal axis almost twice as long as first standardised 4:3 aspect ratio, Edison's 35mm film, (Cossar, 2011, p.28). At the same time, the universally known measure of shot scale, a standing human body and its distance from the lens, still stayed the same despite that the amount of horizontal space varied greatly in a range of aspect ratios (figure 116).

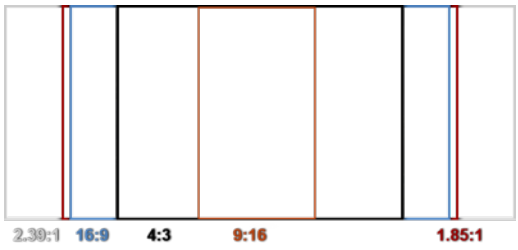

Figure 116: Overview of the most popular aspect ratios. From 9:16 vertical to 2:39 theatrical frame.

To recall details of the shot scale from chapter three, the distance of a standing human body from the camera determines the naming convention for shot scale. For example, the vista shot means a single human body is not distinguishable in the composition; a wide shot encompasses a significant amount of lateral space and headroom above the subject, medium or medium-close up less, while close-up and extreme close-ups were commonly outlined without any headroom. Following that logic, the amount of lateral space in the composition corresponds to the amount of headroom. The exact relationship varies depending on the subject, but in the example of a wide shot seen in figure 118 where the key subject is a leading motorcycle, a protagonist can fit 4-5 times its size on the horizontal axis and 2-3 times on vertical. In contrast, it is not possible to apply a similar rationale to a mise-en-scene housed in a vertical frame since the amount of headroom fits into the pattern of the widescreen shot scale, but that wasn't the case for available horizontal space in the composition. For that reason, the conventions of widescreen shot scale naming might be misleading when a vertical screen is considered as a canvas for moving images. To illustrate the shape difference between various widescreen aspect ratios and a 1:2 tall frame better, I use the same overview of the shot scale introduced in the methodology chapter (figures 13-20, pages 27-28), but this time with the mask covering cropped sections of the image (figures 117-125). As you can see in figures 117-125, all compositions except vista and wide had small or non-existent space on the X-axis.

I left the crop mask at 15 percent transparency to illustrate how much screen area was lost after cropping footage into a 1:2 aspect ratio, since that was one of my main concerns during the re-purposing of widescreen material to a vertical frame.

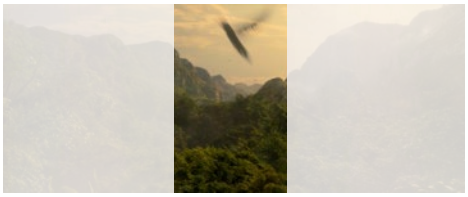
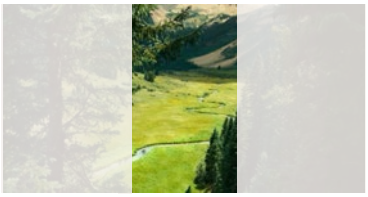

Figure 117: Two version of vista composition.

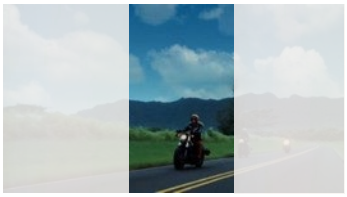

Figure 118: A wide shot.

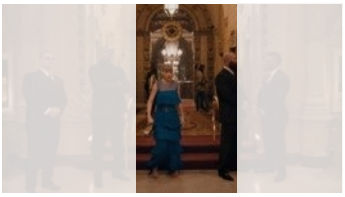


Figure 119: Two versions of full body shot depending on a subject.

Figure 120: Medium shot in 16:9 aspect ratio.

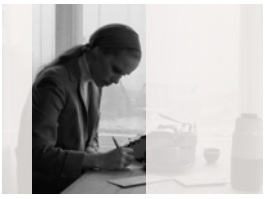


Figure 121: An alternative versions of medium shot in 4:3 aspect ratio.

Figure 122: An alternative versions of medium shot in 2.35:1 aspect ratio.

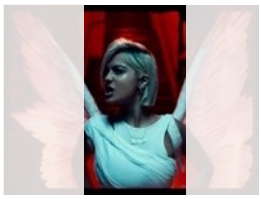

Figure 123: Medium close-up in 4:3 aspect ratio.

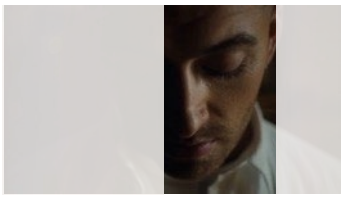

Figure 124: Close up.

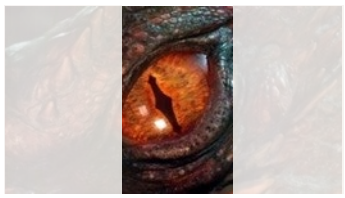

Figure 125: Extreme close-up.

After reassessing the appearance of cropped screenshots, I conclude that the most prominent aesthetic change accompanying a vertical frame was the loss of lateral space next to the subject. The absence of original mise-en-scenes context was exacerbated in vista and wide compositions, less prominent in medium and close-up shots, and minimal in extreme close-up. Consequently, the ability to balance graphic, index, and motion vectors or film multiple subjects having a similar screen prominence become heavily compromised in vertical frame. Good examples of the negotiated screen space in the 1:2 frame are dialogue scenes, where limited lateral area exaggerates any horizontal movements, transforming subtle motion into more intense action as parts of the protagonist's bodies enter or leave the composition. For those reasons, if the storyline commands complex blocking and movements, the camera must be further away from the subject, limiting the shot scale and aesthetic choices. A more defined frame may or may not work for every scene of a particular script, as demonstrated in the key example of this chapter, and if filmmakers embrace the benefits of less ambiguous vertical composition, they also must accept the reduced screen space remaining for the rest of the visual elements.

Chapter Seven: Re-Purposing Loss

With an improved understanding of the principal benefits and limitations of vertical framing learned during previous methodological steps, I searched for a script that was a good fit for the production of artefacts #5 and #6. After exploring the potential of a few other projects, I settled for the short fiction film *Loss* (2019) as the source of suitable cinematic footage. *Loss* is a family drama that follows three protagonists; a woman, a man, and a child in three different life stages. Since most of the action revolves around the mother, the dominance of a single character made this particular script a good fit for vertical re-purposing, as the majority of the shots contained a single subject, occasionally two, and all three in only a few scenes. Furthermore, because I was involved in the making of *Loss* from the scriptwriting stage, I had an excellent understanding of the director's aesthetic vision, as well as insights on how successfully those ideas were translated into images with respect to the production circumstances.

Figure 126: Screenshot of from Loss, before and after colour grading. From 'washed' original files filmed with Log2 gamma curve to the final look.

We shot the film on the South Coast of Wellington with a skeleton crew over one week in a 16:9 aspect ratio. The total running time of the final cut is 10 minutes 33 seconds, split between 58 shots. A principal recording device for this item was the Canon Cinema 300 Mark II camera, with a kit of five Carl Zeiss Compact Prime 2 lenses ranging between 25-105mm (25mmT2.9, 35mmT2.1, 50mm T2.1, 85mm T1.5, 105mm T2). The resolution of the recorded image was full HD 1920 x 1080, with bit-rate of 100mb/s compressed in Apple ProRes HQ[x] codec. Knowing we will be working in the exterior without additional gear to control or modify lighting, I set the image picture profile to a Cine Log2 gamma[x] (figure 126), to get the best possible dynamic range in high luminance lighting conditions, such as a partly cloudy daytime. A 1:2 frame was Netflix's choice for vertical previews, and the requirement to edit the film and a trailer in a tall aspect ratio even slimmer than 9:16 forced me to rethink the composition entirely. The task at hand echoed remarks of Elsie from the chapter five: 'you have to go through your footage with a whole different perspective, you have to start looking at it in terms of that limited space, and the fact it wasn't shot with that in mind' (Elsie). Having Elsie's comment as well as conclusions from chapter four in mind, I considered what elements of the original mise-en-scene could be excluded while keeping the shot relevant for the story's progression.

This is an overview of footage manipulation steps conducted in Adobe Premiere Pro while re-editing the film. Once the final cut of the widescreen film version was locked, I transferred the 16:9 timeline into a 1:2 aspect ratio sequence set for 1080 x 2160 resolution resembling the aspect ratio of Netflix previews. If camera movement was necessary to keep the action within the screen boundaries after punch-in, I simulated corrective camera pans through manipulation of image motion keyframe in the Adobe Premiere Pro effects panel. To further improve the quality of footage, I applied an optical flow[x] filter to compensate for the missing frames created by motion interpolation in the footage and make artificial movements softer. Lastly, if there was no other re-framing option, such as the case of shot #32, I split the original widescreen footage into two pieces and edited them as separate shots. To accommodate the original 1920 x 1080 recording into a 1080 x 2160 frame, the footage needed to be magnified 2.5 times to fill the entire frame.

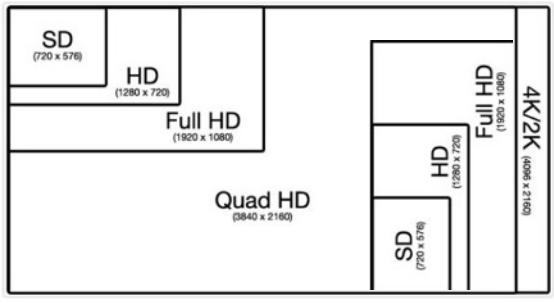

Figure 127: An overview of resolutions from standard definition to 4K.

I was aware that digital re-sizing could affect the image quality, but at the time of the recording, due to the budget constraints and the nature of the remote filming location, the *Loss* (2019) production couldn't afford to film 4K or higher resolution

footage, (figure 127). That was the principal reason why image needed to be upscaled, but as a creative method that was nothing new. In an interview Walter Murch gave to Rodney Hill, Murch discusses the use of image blow-up as a story-driving element in great detail, and even more intriguing, as an editing method he used at times when analogue film was in the midst of remediation with emerging professional digital equipment but mobile recording technology was still far behind in terms of resolution. After describing multiple technical issues of a birds-eye camera rigging in *Youth Without Youth* (2007) and the inability of crew to hit correct focus marks, Murch describes how 'digital zoom' in software eventually produced the desired effect:

> so you start with a wide shot and end up with a close-up; and I think instead of twenty percent, I was enlarging it a hundred and twenty percent. After using it for the first time, I started using that a lot, being very free in the film with blowing up, enlarging the image, which allowed me to accelerate certain cuts (Hill, 2017, p.56).

Following a similar logic, despite resorting to image blow-up, I was still confident the files would not exhibit any significant compression artefacts once packed with H264 codec in an MP4[x] container, one of the most advanced algorithms developed for online distribution of video content. To assess the quality of the footage after post-production manipulation and confirm the render did not create any compression artefacts in the image, I watched the same file on six different mobile screens borrowed from family and friends. Here is the list of the devices, with their respective year of release: iPhone 6 (2014), iPhone X (2017), Samsung Galaxy S9 (2018), Huawei Y5 (2019) and iPhone 12mini (2020). Listed smartphones feature different screen diagonals and resolutions, but I would argue any footage recorded in a bit-rate equal to or higher than 50mb/s in a large camera body, and exported with the use of advanced compression codec from a professional non-linear editing software such as Adobe Premiere Pro will not exhibit any apparent image artefact once played on a small screen.

Figure 128: Visual imperfections are more dominant in a frame with fewer elements.

After re-editing was finished, the colour-grading treatment was the same for the majority of shots, except #30, #32, #36, #42, #44, #46, and #51. Due to a lack of volumes visible in the widescreen composition of those seven shots, I had to separately adjust the highlights and shadows of a few mise-en-scene elements now more prominent in a vertical frame. I.e. The lack of positive volume made by the bright surface of the road passing behind protagonist, and a dark mass of the rock initially visible in the widescreen, made the highlights on the child's hat more dominant in the shot since no other high-key volumes were present in the frame to balance the composition, (figure 128). Lastly, it is essential to note that the audio design and music score were not finalised at the time of writing. The current audio track in artefacts #5 and #6 is a combination of ADR (automated dialogue replacement) and location wild-track[x]. This wasn't an ideal situation as sound design can significantly enhance the prominence of on-screen and unseen off-screen elements in diegetic space, but all edit points, as the most vital part in the building of diegetic space, were already decided.

The all-encompassing rule that guided my rationale during the cropping process was the desire to stay true to the atmosphere described in the film script. While looking at the well-known footage with a fresh pair of eyes and vertical composition in mind, I tested numerous versions of the footage cropping, looking to preserve the scene context, emphasize the actors' emotions, and make sure all key story driving elements stayed within the visible frame. That wasn't a simple task, and despite my best efforts, some compositions exhibited major transformations in the process. As a reference, I also included a link to the original widescreen cut of the film.

- Reference / widescreen version of *Loss*
- Final / artefact #5 vertical version of *Loss*

Figure 129: Loss shot #1.

The following section of this chapter includes a selection of scenes, illustrated by screenshots, that best represents the summary of my rationale for editing decisions I made during the re-purposing process. I occasionally included multiple screenshots of the same scene if necessary to better understand the action.

Envisioned as a static establishing wide composition, the opening shot #1 needed a simulated camera pan to track the subjects until they leave the frame at the left edge (figure 129). The purpose of the movement was to keep the action visible without giving any of the protagonists extra screen time.

Figure 130: Loss shot #12.

Shot #12 was also re-framed to keep the action within the screen boundaries. If I were originally planning shot #12 for a vertical screen, I would suggest more subtle blocking to the director and a shorter running time. In the widescreen version, the audience sees an extended two-step movement of a child that couldn't be replicated due to the lack of side space in vertical version (figure 130). In addition, while doing the same action, the child left the vertical frame quicker than in the widescreen composition, making the movement itself more sudden than intended. Irrelevant features of the image quickly exhaust audience attention, consequently rendering a third of the original widescreen running time redundant as the camera lingers on an out-of-focus background.

Figure 131: Loss shot #13.

Next in the sequence, the widescreen shot #13 is supposed to match the jumping action from shot #12, but once the child leaps above the camera into vertical shot #13 and starts running towards the mother, the audience sees two bodies in the shot with no extra screen space (figure 131). Such hermetic composition tied with a strong motion vector induced by running, accentuated the emotional connection between the two featured protagonists. But, at the same time, part of the running action force originated by flying pebbles was lost in vertical frame. In summary, as a result of re-framing, shot #13 lost the energy of lateral motion vectors, but gained in the emotional intensity.

Figure 132: Loss shot #18.

Shot #18 was part of a dialogue scene where a discreet rack of focus shifts the attention between the subjects during the conversation (figure 132). It is one of the key moments in the film as character relationships become clearer, and is an excellent example of how changes of aspect ratio have a multifaceted effect on dialogue scenes. The emotions on the mother's face together with the subtle body language made already strong directional vectors produced by her gaze become even more prominent in vertical composition. Furthermore, original widescreen blocking included negative space for the child (sky) and positive volume behind the mother (part of the hill), anchoring her body as a very stable foreground object in the composition. A negative space formed by an empty sky in the 16:9 version suggests that the child is absorbing the words, but at the same time scrutinises the mother's monologue to some extent. In cinematic language, a large leading room in front of the child, such as empty third of the screen on the left hand side, indicates there is a space to think, a space for some 'other thoughts'. In contrast, the vertical composition features significantly limited negative space, reduced from approximately 40 percent in widescreen to roughly ten, consequently producing a feeling the child is 'covered' by the mother. Equally important, due to its position, the fraction of negative space left in vertical shot now became associated with the mother, and in this scenario her spoken words and blocking becomes even more prominent. In summary, the vertical composition allowed the audience to see mostly the mother's body in the foreground, and even if the child occasionally reacted, the gesture becomes extinguished by uneven visual prominence. At this point in the story, the subtext induced by vertical composition of shot #18 was welcomed, as the new configuration carried a stronger feeling of the mother's dominance over the young child.

Figure 133: Loss shot #19.

As a counterpoint to the previous example, shot #19 of the dog looking up at his master lost most of its context after re-cropping, since the audience cannot see where the dog is looking in a vertical 1:2 frame (figure 133). Because shot #19 was visualised as a brief reaction, and the director consider the relationship between the man and the dog are clear at this point, we never filmed an alternative version, which became an issue in the restructuring of that scene in a vertical frame. If the time and budget allowed, I would concur with comments Claire made regarding filming for Quibi in chapter five, and make an effort to record a separate take with only vertical composition in mind.

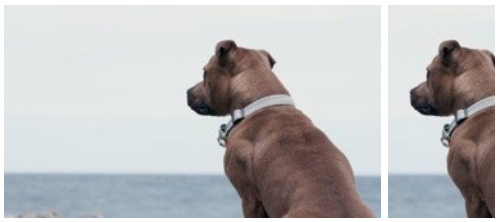

Figure 134: Loss shot #21.

Shot #21 is another good example of how re-arrangement of the composition in distinct aspect ratios can introduce ambiguity into the original idea. In the widescreen version, an extended index vector of the dog combined with the negative space of open sea dominated this minimalistic composition (figure 134). Desired audience interpretations of such composition was a breathing space, or a free space that can be occupied by new ideas and hope. In contrast, once housed in a 1:2 aspect ratio, a similar mise-en-scene of a 'trembling' dog facing a very short horizon created an altered connotative meaning, one that didn't align with the idea from the original script. What was wide breathing space in front of the dog in widescreen shot, now becomes confined and hermetic due to a much smaller 'nose' room. Furthermore, most of the dog's back, where the trembling motion was visually prominent, was not visible in the newly formed vertical composition, encoding a different emotion from the animal due to a lack of movement.

Figure 135: Loss shot #23.

The following scene illustrated by shot #23 is arguably the best example of how cropping of the original frame formed a different dynamic in the composition and introduced a subtle but nonetheless important subtext. The principal difference between the two mise-en-scenes derives from the amount of time we see protagonists on screen. In widescreen, the entire family of three is visible for 80 percent of the scene's running time as they walk along the road (figure 135). The blocking and position of their bodies in space, together with the connotative meaning of positive and negative volumes, reflect the atmosphere of disconnect in this family, but a similar action housed in vertical frame projects an even stronger sentiment of disconnect because the audience cannot see the entire group on-screen together at any point. Due to a slimmer frame, the protagonists alternate on the screen, so at the beginning of the sequence only the daughter and father share screen space, followed by only father and mother in the composition.

At this point in the *Loss* story, staging of the father in the middle of widescreen composition was highly motivated, but after re-framing his character got twice as many on screen seconds in vertical shot compared to the daughter and mother, whereas split from 'family of three' into 'two couples' misleadingly inferred his role as a unifying character in the scene. Furthermore, his central blocking in combination with 1:2 aspect ratio also changes the notion of the gestures exchanged between him and the mother. In the second half of the shot #23 running time, when the father turns around towards the mother she looks away. At that moment in widescreen, his body is placed at approximately the middle of the shot, while space on the left hand side was just a moment before occupied by his daughter. If a similar exchange is observed in vertical frame, the body of a father, thus his presence, is already 'half way out of the frame' when he turns around, and at that moment the daughter left the screen a while ago in terms of screen time. Because his presence in the scene is 'cut in half' by the time he looks at the mother, her disapproving reaction is even stronger because they are 'miles apart' from each other with respect to available horizontal screen space. Lastly, where her reaction in widescreen can be coded as defiant with an abundance of negative volume formed by the sky, having her face close to the edge of the frame in vertical composition when she delivers the reaction suggests a lack of options and alters the power balance between the characters.

Figure 136: Loss shot #32.

Shot #32 was scripted as a symmetrical 16:9 M composition of a dialogue between the mother and (now older) daughter with no camera movements (figure 136). An equal prominence of both characters in the composition, and lack of secondary motion meant the focus of viewer attention should be on the snappy comments of the mother and subtle reactions of her daughter. But as I analysed variations of shot #32 after cropping, I could not find the composition able to house both characters on screen simultaneously. Considering the shot's length, a simulated pan between two protagonists would be too fast, thus unjustifiably mismatching the pace of the mother's voice, therefore the only other available option to keep both characters on the screen at the appropriate time was to split the shot into two parts. In doing so, the audience could see the mother's dialogue delivery and daughter's reaction while the camera stayed stationary, honouring the original guidelines from the script. Despite having a solution that gave an approximately equal amount of screen time to both characters, I concluded the atmosphere between mother and daughter could not be perceived in the same way in the newly formed vertical composition, as they did not share the same screen space anymore. By giving them separate screen

time, the scene lost part of the emotional exchange visible in the two-shot composition, and the characters look more separated than they should be at this point in the story.

Figure 137: Loss shot #37.

The last sample included in this selection is the insert of shot #37, as an example of widescreen reframed composition that neither benefited, lost energy nor gained additional subtext from tighter framing. As the mother frantically searches the rock pools, her wide movements occasionally extend into the off-screen space, but contrasting the previous examples both compositions were justified in conveying the feeling of despair (figure 137).

Having the vertical version of the film finished, I could proceed to the making of artefact #6. At this stage of the research, the editing of the trailer was a straightforward process for me because I did not have to start with complete footage revision again. With sound knowledge of what shots are most effective in a slimmer frame, I could make a selection quickly and with great confidence (figure 138).

Figure 138: A selection of screen-shots from vertical trailer for Loss.

In a skilfully carved visual sequence, there are no redundant frames. That is, a half a second longer shot of protagonist reaction before leaving the frame can suggest hesitance, or in a different scenario, if a similar scene was cut before showing the intention to leave, the connotative meaning of the sequence will change. The need for precision is even more apparent when editing trailers because the sequence of shots needs to digest the emotions and atmosphere of the much longer primary piece with a limited number of shots. Being a melodrama, I deemed that a slower editing rhythm was a good fit for *Loss*, but because the movie was short, I had fewer scenes to present to the audience than the editor of a full length feature. Most of the 12 shots I selected were M, with a few CU's and a single L static image at the beginning and end, known as poster frames or thumbnails in the video. The selection of shots also corresponded with the observed trend that the vertical trailers have more M, MCU and CU shots than their widescreen counterparts described in chapter four. Looking to re-create the footage aesthetics observed in Netflix trailers, I also placed a gradient mask to darken the lower third of the shot and made closed captions. In the case of Loss, the gradient mask did not obscure any key story-driving elements in the mise-en-scene, but it did affect image luminance levels, rousing darker subtext than originally envisioned. Lastly, adhering to the dominance of camera movements along the Z axis in online vertical content aimed at mobile screen, I introduced a subtle digital zoom as a substitute for the lack of dolly movements to induce a forward momentum in otherwise static composition. Here is the link to the last thesis artefact.

- Artefact #6 vertical 1:2 trailer for Loss

In summary, after cropping, the film's vertical version exhibited several significant changes. Due to the fact that fewer elements were visible on the screen, some scenes projected a stronger emotion, and the vertical version of Loss (2019) rendered a stronger sense of claustrophobia, a sentiment that wasn't in the original script. Multiple landscape compositions, such as wide shots of protagonists on a coastline battered by waves, were designed to indicate a sense of loneliness but not claustrophobia, and the lack

of space on the horizontal axis meant several essential storytelling elements couldn't be present for the majority of running time. Consequently, without the prominent visual presence of coastline and waves in the composition, the context of a location was diminished, as well as the role of 'the Ocean' as a character in the story. Lastly, despite having a slower editing pace than blockbuster previews I analysed in chapter four, artefact #6, a vertical trailer I edited for Loss, accurately conveyed the atmosphere from the vertical film, artefact #5.

Chapter Seven: Re-Purposing 16:9 Footage to 1:2 Aspect Ratio – Conclusion

The editor's ability to manipulate the image aspect ratio is a versatile storytelling tool, but it should be used with great care as I learned from making artefacts #5 and #6. Some *Loss* (2019) script elements appear strengthened when converted to a 1:2 aspect ratio, while others lost prominence due to weaker on-screen contextual links. For those reasons, the two versions of *Loss* are similar enough to be perceived as two accounts of the same story, but the vertical aspect ratio brought additional subtext in multiple scenes which cumulatively changed the balance between the characters and altered the atmosphere envisioned in the script. Equally important, the editing of the two-shot dialogue scene required more cuts due to insufficient space for both characters, and tight framing often prompted the cropping of large parts of the protagonist's bodies. Because the newly formed edges of vertical diegetic space often cut actor bodies along unorthodox lines (for example, part of the head in a medium shot), and reformed the balance of positive and negative volumes in the composition, I struggled to honour character dynamics for some of the key scenes designated in the script.

When I reflected on the process of footage assessment and reviewed my notes on the highlights selected for the trailer, I recalled spending most of the time debating my re-purposing choices of shots with multiple protagonist. Having the addition of a dog as a fourth character, the footage from the beginning of the movie was the most challenging to re-mould into a vertical frame. During the last third of the film, when the story mostly revolved around two characters or only the mother, it was much simpler to determine the key focus of attention and decide what kind of re-framing would be the most beneficial for the building of vertical diegetic story space.

Recalling comments from the interviews conducted in chapter five, if the footage was filmed with both a vertical and horizontal frame on the director's mind, we would pre-plan a dedicated vertical pass for pivotal moments such as shots #19, #23 and #32, to make sure those scenes do not lose correct, or provoke incorrect context for character relationship. Furthermore, if production of *Loss* discussed dissemination on mobile and planned a vertical version of the film, before commencing the vertical pass, the actors would get prior instructions to limit the lateral motion and blocking would be arrange mostly along the Z axis.

To summarise, while some vista and wide shots suffered from the lack of location context, others such as dialogue scenes benefited as a slimmer frame emphasised the emotion and highlighted subtle reactions during the conversations. A more narrow canvas also worked well for this particular story by strengthening the feeling of hermetic space, that in turn underlined limited options for the leading character. Suppose a spectator or media professionals I interviewed in chapter five approaches viewing experience of the alternative version of Loss with no prejudice on aspect ratio; I am optimistic one would be equally immersed in vertical versions of Loss as it would be in the original widescreen because the emotion in both stayed true to the script.

Chapter Eight: Thesis Conclusion and Reflection Statement

The salient features of the vertical frame commonly coded in the audience's perception of the vertical aspect ratio such as intimacy and authenticity, were verified throughout text analysis in chapter four, comments of professionals in chapter five, as well as the creation of thesis artefacts in chapters six and seven. For that reason, if a professional media creator works on a project aimed at a mobile audience, a simpler mise-en-scene in combination with a vertical frame carries that feeling stronger than widescreen. From the variety of texts I analysed and six created artefacts, I learned that simplistic compositions with a single dominant point of interest, such as a medium shot of the human body, benefit the most from a more defined vertical composition. Correspondingly, vista and very wide shots featuring complex mise-en-scene and multiple points of interest were not a good fit for the slimmer frame, since such compositions often lost some of the original contexts after re-framing.

Since there aren't many options to house a vertical frame in widescreen other than image blow up or pillar boxing, which undoubtedly deteriorates image quality and perception of content, I would argue that the principal strength of the vertical frame resides with the development of the original content. The analysis revealed, and the interviewed professionals agreed, that when a story is scripted with a tall frame and a mobile screen in mind, the engagement potential of vertical video cannot be denied. Equally important, a lower production value coded in audience perception of single take vertical videos may not be fit for elaborate sets, complex mise-en-scene and camera movements, but in return, such videos are much cheaper to produce. With that in mind, it becomes clearer why the widescreen music videos are often the leading ones, and vertical versions are used as an engagement booster targeting mobile audiences. Although single take vertical music video developed a unique repetitive dolly movement aimed to improve attention synchrony of static compositions, observed movements were non-existent in the larger sample, confirming such aesthetic did not become widely popular.

The interviews in chapter five revealed professionals' attitudes towards vertical video did change to the point it cannot be regarded as backlash anymore, but the resistance to its full acceptance as a cinematic expressive format still exists. I would conclude the antagonism resides in the lack of established workflows that practitioners can follow as standard. For the same reasons, media professionals are reluctant to accept a smartphone as a primary recording device fully because there is no set of technical and aesthetic criteria's that guarantees the footage quality it can deliver. However, despite not being willing to accept vertical screens as a primary outlet for the distribution of their content, professionals are willing to harvest the engagement potential of vertical video cut-downs on social media, and learn how to make an informed selection of aspect ratios targeting the audience on a specific online platform.

While filming original artefacts in chapter six, the mobility of a smartphone was a great advantage, allowing effortless camera movements and cutting down the set-up time. However, despite being able to record footage in a resolution and bit-rate adequate for a professional workflow, aesthetics of deep focus and lack of precision focus controls prove to be two critical elements that limited my creativity while shooting with a smartphone. Moreover, the absence of high-quality optical zoom, or the ability to quickly change the battery and duplicate media to a workstation are additional technical limitations that question the choice of smartphones as a principal recording device in a professional workflow. As anticipated through the analysis in chapter four, filming a music video of a solo performer was a better fit for a vertical frame than the set of a multipiece band, where an abundance of action and visual accent compromised my ability to find a balanced frame and emphasize individual performers.

Chapters seven demonstrated that if a final segment is aimed for distribution in multiple aspect ratios and online platforms, the creator has to be mindful that the widescreen composition cannot always be adapted to a vertical frame. By cropping the footage to a different aspect ratio, the editor creates a new edge of the frame, one that is not envisioned or justified by script, thus re-designing the original mise-en-scene. For those scenes that render a significantly altered meaning, a production should budget the

time to re-organise mise-en-scene along the Z-axis, change the camera angle and update blocking instructions for onscreen talents to ensure key story driving elements or context of the scene won't be lost in a slimmer frame. Moreover, if the author has only a mobile audience in their mind, a simplified mise-en-scene and shot scale will be welcomed, as details of a complex composition can be lost on a small screen.

Reflecting on the research steps, I would conclude that mixed methodology proved to be a good choice in answering multiple facets of the principal thesis question. Nevertheless, for even stronger conclusions in chapter five, I would gather a larger data set by interviewing more media professionals since some of the key themes had overlaps, making it harder to find definitive conclusions in the analysis. Another amendment to the methodology relates to the choice of filming locations in chapter six. In order to avoid the connotative meanings of different luminance levels and aesthetics in the acquired footage, coming from the use of two studios for solo performers and a multipiece band, I would make sure all thesis artefacts were filmed in the same space. For the analysis of the re-purposing process in chapter seven, I would ideally have a feature length film as a key example, preferably one that I wasn't part of production, because that fact would make my editing decisions even more objective and closer to the working conditions of editors employed by Netflix.

Additionally, although parts of this research, particularly chapter five, might be pertinent to the inquiries into production cultures, I decided not to dive into that field to keep the stronger focus on participants' technical and aesthetic decisions during the production of vertical content. While I recognise the significance of studying production cultures, I also acknowledge the intricacies commonly associated with such research. A deeper investigation into the management structure, budgets and other production circumstances are important facets of any professional workflow, and for a better understanding of them, this research can be supplemented with interviews from a larger variety of geographical locations and a bigger data set sampled from the millennial generation, which would more accurately reflect the contemporary demographics of professionals employed in the media industry. As an alternative approach, fieldwork would also be a beneficial method for a production cultures approach to the topic, involving participant observation in different settings ranging from studio to home office and deeper research into the notion of what is accepted as a professional filming environment. A good examples of such research can be found in Vicki Mayer's, Miranda Banks and John Caldwell's *Production Studies: Cultural Studies of Media Industries* (2009) as well as in the follow-up study edited by the same group encompassing social media, *Production Studies, The Sequel!* (2015).

Equally important, despite the fact I did not inquire into the matter of representation in the research, I pay close attention to the new developments in the field and have observed that vertical video has become more prominent online during the past four years; thus, its aesthetics are more relevant in a variety of contexts. Simultaneously, the increase in production and consumption of vertical content, spearheaded by the global popularity of TikTok, has put a fresh spotlight on the relevance of vertical videos in studies on personal representation online. For that reason, the conclusions from this study might support researchers interested in a deeper analysis of what kind of topics users choose to tell within a vertical canvas or what genre of scripts media professionals develop for tall format. It would also be beneficial to analyse what topics dominate celebrity-made vertical videos in contrast to user-generated vertical content, as both groups look to benefit from the sense of honesty commonly associated with the vertical frame. Looking even deeper into implications of vertical filming, follow up studies can inquire how will this new format and its popularity, shape a focus on the isolated individual performer as opposed to a group, and what does this mean culturally? In addition, a separate study can probe what sort of implications does the vertical frame have for the representation of youth, or the relationship between media representation and body image.

Furthermore, the growing ubiquity of vertical video opens new questions in the field of technologies, and distribution channels. The fact that a well-established platform such as Instagram amended its posting guidelines in July 2022, favouring vertical content in response to its popularity on TikTok, is a prime example of video remediation and a motive to continue my research into the specifics of vertical video evolution associated with a particular online platform. These new forms of representation also

have an immense impact on the growth of celebrity culture in the online environment and the average user's drive towards individualism as a person replicating the aesthetics of the most popular online texts. Another cultural consequence of vertical video proliferation in the online environment is exemplified in the tendency that professional content creation workflows shift towards multitasking personally instead of sub-specialisation, that in turn, changes the collaborative practices and affects the selection of topics and aesthetics of produced footage.

Lastly, I have to acknowledge how drastically my thinking on vertical video had changed since I started this research from the days when I worked as a full-time media professional. Once I had the opportunity to teach and spend more time with the younger generation of content creators, I realised the vertical frame is not better nor inferior to widescreen, but part of expressive practice with a limited application. If used in combination with the adequate script, given enough pre-production time and disseminated on the mobile platform, it will often outperform widescreen, particularly regarding the return on investment against the initial production cost. A vertical content with brief running times aimed for the mobile screen is the type of production where the use of a smartphone as a primary recording device becomes incentivised, despite its evident inferiority when contested with ergonomics and footage quality recorded with DSLRs, SLRs, or larger camera systems. Ultimately, for versatile and open-minded media professionals who embrace remediation of formats, the choice of vertical against widescreen aspect ratio or smartphone versus professional camera body becomes a question of choosing the right tool for the job, since the aesthetic capabilities of the vertical frame are without a doubt different from traditional widescreen, but the trick is in knowing when to use it.

Bibliography

Allain, R. (2014). *A Closer Look at Slow Motion Video on the iPhone 6*. Wired.

Arnheim, R. (1974). *Art and Visual Perception, Second Edition : A Psychology of the Creative Eye.* University of California Press.

Arnheim, R. (1982). *The Power of the Center*. University of California Press.

Banks, M., Conor, B. & Mayer, V. (2015). *Production Studies, The Sequel!: Cultural Studies of Global Media Industries*. Routledge

BBC. (2021). *Nearly a third of waking hours spent on TV and streaming.* BBC News.

BBC. (2021). Subtitle Guidelines. BBC.

Benjamin, W. (1935). *The Work of Art In the Mechanical Age of Reproduction.* Penguin Books.

Bereznak, A. (2019). *This Music Video Has Been Modified From Its Original Version (and Now It's Vertical).* The Ringer.

Bolter, J. & Grusin, R. (1998). *Remediation - Understanding New Media*. MIT Press.

Bordwell, D. & Thompson, K. (1985). *The Classical Hollywood Cinema/Film Style & Mode of Production to 1960*. Routledge.

Bordwell, D. & Thompson, K. (2001). *Film Art: An Introduction (Vol. 6th)*. New York, USA: McGraw Hill.

Bordwell, D. & Thompson, K. (2009). *Film Art: An Introduction Film, 9th edition*. McGraw-Hill.

Braun, V. & Clarke, V. (2006). *Handbook of Research Methods in Psychology: Vol. 2.* The American Psychological Association.

Britannica. (2019). *Cinématographe film technology*. Britannica.com

Brown, B. (2016). *Cinematography: Theory and Practice: Image Making for Cinematographers and Directors 3rd Edition*. Routledge.

Caldwell, J. T. (2008). *Production Culture: Industrial Reflexivity and Critical Practice in Film and Television*. Duke University Press Books.

Canella, G. (2018). Video Goes Vertical- Local News Videographers Discuss the Problems and Potential of Vertical Video. *Electronic News*, *12*(2) 75-93.

Ceci, L. (2022). *Mobile internet usage worldwide - statistics & facts*. Statist.com.

Cirisano, T. (2019). *How the Vertical Music Video Finally Took Off*. Billboard.com.

Cisco. (2020). *Annual Internet Report – White Paper (2018–2023)*. Cisco.

Cossar, H. (2007). *Snakes and Funerals: Aesthetics and American Widescreen Films.* Georgia State University.

Cossar, H. (2011). *Letterboxed: The Evolution of Widescreen Cinema*. The University Press of Kentucky.

Cooper, P. (2019). *How Does the YouTube Algorithm Work? A Guide to Getting More Views*. Hootsuite.

Deloitte. (2017). *The kids are alright: no tipping point in TV viewing trends for 18-24 year old's* Deloitte Insights.

Deloitte. (2019). *Tech Trends 2019*. Deloitte Insights.

Dempsey, T. (2013). *Compare digital camera sensor sizes: 1"-Type, 4/3, APS-C, full frame 35mm*. PhotoSeek.com.

Dmytryk, E. (1986). *On Filmmaking*. Focal Press.

Dyer, R., Kaplan, A., Willemen, P., & Hill, J. (1988). *The Oxford Guide to Film Studies*. Oxford University Press

Filmic Pro. (2020). *Bring Your Vision to Life*. FiLMiC Inc.

Fischer, B., & Ramsperger, E. (1984). *Human express saccades: extremely short reaction times of goal directed eye movements*. Experimental Brain Research, 57, 191-195.

Friedberg, A. (2009). *The Virtual Window From Alberti to Microsoft*. MIT Press.

Gibson, J. (1979). *The Ecological Approach to Visual Perception.* Routledge.

Glove & Boots. (2012), *Vertical Video Syndrome – A PSA.* YouTube.

Glove & Boots. (2019). *This is Vinny and Damien, the humans behind Glove and Boots*. Facebook.

Grabyo. (2020). *Grabyo's At Home Video Trends 2020: Lockdown's lasting effects on the video industry*. Grabyo Press.

Grabyo. (2021). *THE FUTURE OF BROADCAST The cloud platform for live production and editing*. Grabyo Press.

Handley, L. (2019). *Nearly three quarters of the world will use just their smartphones to access the internet by 2025.* CNBC.

Hardawar, D. (2020). *Quibi's secret weapon: Videos that work in portrait and landscape mode*. Engadget.

Harmon, L. (2020). *Has the TikTok Generation Actually Changed the Future of Television?* Observer.

Harper, C. (2011). *Letterboxed: The Evolution of Widescreen Cinema*. University Press of Kentucky.

Hausken, L. (2013). *Thinking Media Aesthetics. Media Studies, Film Studies and the Arts*. PL Academic Research.

Hill, R. F. (2017). *A Minute to Learn, a Lifetime to Master: An Interview with Walter Murch* Journal of Film and Video , Vol. 69, No. 3 (Fall 2017), pp. 54-62. University of Illinois Press.

Hill, S. (2013). *From J-Phone to Lumia 1020: A complete history of the camera phone*. Digital Trends.

Hovet, T. (2017). The Persistence of the Rectangle. *Film History, 29*(3), 136–168.

Huawei. (2019). *One-Handed Mode.* Huawei Consumer.

IGTV. (2018). *Welcome to IGTV, our New Video App*. Instagram Press.

ITU-R. (2015). *Parameter values for the HDTV standards for production and international programme exchange*. International Telecommunication Union.

Keep, D. (2017). *Remembering Hiroshima.* Aspera, *Sightlines Journal*, *2*.

Keep, D. (2018). Smartphones and Evocative Documentary Practices. *Mobile Story Making in an Age of Smartphones* (pp.41-50).

Lee, S., Kyung, G. & Yi, J. (2019). Determining Ergonomic Smartphone Forms With High Grip Comfort and Attractive Design. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 61*(1).

Leichtman Research. (2019). *Consumer Spending on Pay-TV is Levelling off*.

Lofland, J. & Lofland, L. H. (1995). *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis.* Wadsworth Publishing.

Lotz, A. D. (2017). *Portals: A Treatise on Internet-Distributed Television.* Michigan Publishing

Manovich, L. (2001). *The Language of New Media*. MIT Press.

Marscelli, V. (2005). *The Five C's of Cinematography: Motion Picture Filming Techniques*, Silman-James Press.

Mayer, V., Banks, J. & Caldwell, J. (2009). *Production Studies: Cultural Studies of Media Industries 1st edition*. Routledge

McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. McGraw-Hill.

Meeker, M. (2015). *Internet Trends 2015*. Bond Capital.

Meeker, M. (2019). *Internet Trends 2019*. Bond Capital.

MINA. (2009). *The Mobile Innovation Network and Association*. MINA.

MPA. (2019). *A comprehensive analysis and survey of the theatrical and home/mobile entertainment market environment for 2019.* Motion Picture Association.

MPA. (2021). *A comprehensive analysis and survey of the theatrical and home/mobile entertainment market environment for 2019.* Motion Picture Association.

MOVR (2018). *Mobile Overview Report October - December 2018*. Scientia Mobile.

MOVR (2019). *Mobile Overview Report October - December 2019*. Scientia Mobile.

MOVR (2020). *Mobile Overview Report October - December 2020.* Scientia Mobile.

MOVR (2021). *Mobile Overview Report October - December 2021.* Scientia Mobile.

Murch, W. (2005). *In the Blink of an Eye.* Silman-James Press.

National Geographic. (2018). *Media Alert: National Geographic Becomes One of the First Media Companies to Launch Long-Form Content on Instagram, Airs Full 'One Strange Rock' Finale*. Nat Geo Press.

Neal, D. & Ross, M. (2014). *Vertical Film 2014*.Walrus & Carpenter Productions LLC.

Neal, D. & Ross, M. (2015). *Vertical 2015.* Walrus & Carpenter Productions LLC.

Neal, D. & Ross, M. (2018). *Mobile Framing: Vertical Videos from User-Generated Content to Corporate Marketing.* Springer International Publishing.

Netflix. (2018). *Introducing Mobile Previews*. Netflix Media Centre.

Netflix. (2020). *A Better Netflix Experience.* Netflix support.

Netflix. (2020). *Cameras and Image Capture.* Netflix Partner Help Centre.

Newman, N., Fletcher, R., David, A. (2016). *Digital News Report 2016.* Oxford: Reuters Institute for the Study of Journalism.

Neuman, W. L. (2011). *Social Research Methods: Qualitative and Quantitative Approaches*. Allyn and Bacon.

Oliver, D. (2021). *'Stan' culture needs to stop – or at least radically change. Here's why*. USA Today.

Ooyala. (2015). Brightcove completes acquisition of Ooyala.

Peterson, N. (2018). *The Rise of Vertical Video: How Top Brands Are Adopting a New Content Format*. Forbes.

Quibi. (2020). *Quick Bites. Big Stories*. Quibi Info.

Robinson, S. (2012). *The Interview: A Process of Qualitative Inquiry/Encyclopaedia of Media Studies.* Wiley-Blackwell.

Raptis, D., Tselios, N., & Kjeldskov, J. (2013). *Does Size Matter? Investigating the Impact of Mobile Phone Screen Size on Users' Perceived Usability, Effectiveness and Efficiency*. Proceedings of the Mobile HCI conference, Munich.

Ross, M. (2014). Vertical Framing: Authenticity and New Aesthetic Practice in Online Videos. *Refractory: A Journal of Entertainment Media*, *24*.

Ross, M. & Glen, M. (2014).'Vertical Cinema: New Digital Possibilities. *Rhizomes: Cultural Studies in Emerging Knowledge, 26*.

Ryan, K. (2017). Vertical video: rupturing the aesthetic paradigm. *Visual Communication*, *17*(2), 245–261.

Salt, B. (2006). *Moving into Pictures: More on Film History, Style, and Analysis*. Tadmor Street

Samsung. (2019). *43" Vertical The Sero 4K Smart TV 2020*. Samsung Press.

Schleser, M., & Berry, M. (Eds.). (2018). *Mobile Story Making in an Age of Smartphones*. Palgrave Pivot.

Smith, T. (2011). *The Attentional Theory of Cinematic Continuity.* Birkbeck, University of London.

Smith, T. (2013). *Watching You Watch Movies: Using Eye Tracking to Inform Cognitive Film Theory.* Birkbeck, University of London.

Snapchat. (2020). *Snap INC news*. Snap INC Press.

Verizon. (2018). *Video content: What millennials want*. Verizon media.

VEVO. (2020). *The World's Leading Music Video Network*. VEVO Press.

VFF. (2014). *Vertical Film Festival*. VFF Archive.

Viola, B. (2020). *Ascension, The Road to St. Pauls*. Bill Viola.com.

Walzer, D. (2017). *Independent music production: how individuality, technology and creative entrepreneurship influence contemporary music industry practices*. Creative Industries Journal, 10:1, 21-39.

Warner. (2016). *Vevo and Warner Music Group Announce Partnership*. Warner Music Group/News.

Willis, J. W. (2007). *Foundations of Qualitative Research: Interpretive and Critical Approaches*. Thousand Oaks.

YouTube VEVO. (2020). *What is VEVO?*. YouTube/News.

YouTube. (2020). *YouTube by the numbers*. YouTube Press.

Zettl, H. (2016). *Sight Sound Motion: Applied Media Aesthetics* (8th ed.). Thomson Wadsworth

Appendix 1 – List of mobile devices

(Link to Excel sheet in Google Drive)

Appendix 2 – List of all texts from YouTube and Netflix considered for the analysis

(Link to Excel sheet in Google Drive)

Appendix 3 – List of interview questions

Q1: Please state your name and the current professional title for the record?
Q2: What is your background and how many years have you worked in media industries (including TV, film, social media and OTT)?
Q3: What is your educational background?
Q4: Can you please provide more details on your current role?
Q5A: Concerning your current role, how much time did you usually spend in pre-production, field production and post-production?
Q5B: When and what was your first contact with vertical content?
Q6: What was your first project aimed to be delivered in a vertical frame for an online distribution outlet?
Q7: Was the content distributed on more than one online channel?
Q8: What was the length of the final delivered video?
Q9: What was a primary recording device? Were there any B cameras involved? If yes – was it the same or different device? (E.g. smartphone, DSLR or large camera body)
Q10: What was the size of the crew assigned to the project and turnaround time?
Q11: What was the budget of the project?
Q12: In your opinion, how does the choice of the recording device affect aesthetic decisions?
Q13: Following up the previous question, how does the choice of the recording device affect technical decisions?
Q14: How does the choice of distribution platform affect aesthetic decisions?
Q15: On the same note, how does the choice of distribution platform affect technical decisions?
Q16: In your opinion, does the aesthetics of vertical video impose any significant workflow changes in comparison to traditional widescreen production? Or what are the main attributes of a vertical frame?
Q17: If not mentioned earlier - Did you ever used a smartphone as a primary recording device?
Q18: If yes - Is there any difference in the set operations when the primary recording device is a smartphone against a larger camera body?
Q19: If no - How would you prepare for the use of a smartphone as a primary recording device versus production and post-production workflow dictated by a large camera? E.g. Compare a smartphone as an extremely portable system able to capture audio and video while being operated by a single person with ARRI Alexa, RED Dragon, or similar professional system.
Q20: Going back to aesthetics. Once the footage is on the timeline in non-linear post-production software, what is the primary criteria for the assessment of the take's quality?
Q21: Follow up if necessary – would you highlight any differences in criteria assessing the quality of widescreen and vertical footage?
Q22: What video editing software do you and your colleagues commonly use?
Q23: Were you ever tasked to re-purpose widescreen archival footage to the vertical frame?
Q24: If yes, can you elaborate on your thought process? E.g. what are the first key decisions that come on your mind when you are assigned to re-purpose the widescreen footage to a vertical frame?

Q25: At the time you worked on your vertical project, were there any templates or other specialised methods in video editing software tuned for vertical content?
Q26: Are you aware of the application Quibi?
Q27: If yes - What are crucial mise-en-scene elements that authors need to consider while producing the piece for distribution outlet such as Quibi (which screens vertical and horizontal aspect ratio simultaneously)?
Q28: If no – How would you prepare for the production when final delivery format is a hybrid between widescreen and vertical frame?
Q29: Online distribution platforms often suggest immediacy. In your opinion, how has immediacy influenced production and post-production decisions?
Q30: Is there anything else you would like to add that I didn't ask you?

Appendix 4 – Spreadsheet with detailed breakdown of dominant 2D and 3D characteristics with key widescreen samples next to vertical videos

(Link to Excel sheet in Google Drive)

Note: All viewing numbers for music videos referenced in this document were last time updated in May 2022.

Appendix 5 – Summary of dominant 2D and 3D characteristics with highlights

Music Videos		A: magnetism & asymmetry of the frame	B: dominant directional vectors	C: objects figure/ground relationship	D: Z-axis articulation	E: secondary framing	F: direct address method
Nothing Breaks Like a	W	Weak	Strong	Strong	Detectable	Strong	Strong
Nothing Breaks Like a	V	Weak	Detectable	Strong	Strong	Strong	Strong
Last Hurrah	W	Weak	Strong	Strong	Detectable	Strong	Strong
Last Hurrah	V	Weak	Strong	Strong	Strong	Strong	Strong
You Should See	W	Weak	Strong	Weak	Weak	Weak	Weak
You Should See	V	Detectable	Detectable	Strong	Weak	Weak	Detectable
Delicate	W	Weak	Strong	Strong	Strong	Strong	Weak
Delicate	V	Detectable	Strong	Strong	Detectable	Weak	Strong
I Don't Wanna be You	V	Strong	Strong	Weak	Detectable	Strong	Weak
Camila C. - Havana	W	Detectable	Strong	Weak	Strong	Strong	Weak
Camila C. - Havana	V	Detectable	Strong	Detectable	Detectable	Strong	Detectable
Harleys In Hawaii	W	Detectable	Strong	Detectable	Detectable	Weak	Weak
Harleys In Hawaii	V	Weak	Strong	Detectable	Detectable	Detectable	Detectable
Nicki M. - Chun-Li	W	Weak	Strong	Weak	Strong	Strong	Strong
Nicki M. - Chun-Li	V	Detectable	Strong	Weak	Strong	Weak	Strong
How Do You Sleep?	W	Weak	Strong	Strong	Strong	Detectable	Strong
How Do You Sleep?	V	Weak	Strong	Strong	Strong	Weak	Strong
Too Good At Goodbyes	W	Weak	Strong	Strong	Strong	Strong	Detectable
Too Good At Goodbyes	V	Weak	Strong	Strong	Detectable	Weak	Strong
Halsey - Without Me	W	Detectable	Strong	Strong	Detectable	Detectable	Weak
Halsey - Without Me	V	Weak	Strong	Strong	Detectable	Weak	Strong

Figure 139: Alternative arrangement of figure 37 from chapter four

Trailers		A: magnetism & asymmetry of the frame	B: dominant directional vectors	C: objects figure/ground relationship	D: Z-axis articulation	E: secondary framing	F: direct address method
F1 - Drive to Survive	W	Detectable	Strong	Strong	Detectable	Detectable	Weak
F1 - Drive to Survive	V	Detectable	Strong	Strong	Detectable	Weak	Weak
Glow Up	W	Detectable	Detectable	Strong	Weak	Detectable	Detectable
Glow Up	V	Weak	Detectable	Strong	Weak	Weak	Detectable
Inside Bill's Brain	W	Detectable	Detectable	Strong	Detectable	Detectable	Weak
Inside Bill's Brain	V	Detectable	Detectable	Strong	Detectable	Detectable	Weak
Marriage Story	W	Detectable	Detectable	Strong	Weak	Detectable	Weak
Marriage Story	V	Detectable	Detectable	Strong	Detectable	Detectable	Weak
Mowgli	W	Detectable	Strong	Detectable	Strong	Detectable	Weak
Mowgli	V	Detectable	Strong	Strong	Detectable	Weak	Weak
Orange is the New Black	W	Detectable	Detectable	Strong	Detectable	Detectable	Weak
Orange is the New Black	V	Detectable	Detectable	Strong	Weak	Weak	Weak
Patriot Act	W	Detectable	Detectable	Strong	Detectable	Weak	Weak
Patriot Act	V	Weak	Weak	Strong	Weak	Weak	Strong
Stranger Things	W	Detectable	Strong	Strong	Strong	Strong	Weak
Stranger Things	V	Detectable	Detectable	Strong	Detectable	Detectable	Weak
Sugar Rush	W	Detectable	Detectable	Strong	Detectable	Weak	Weak
Sugar Rush	V	Detectable	Strong	Strong	Detectable	Weak	Weak
The Ballad of	W	Detectable	Detectable	Strong	Strong	Detectable	Weak
The Ballad of	V	Detectable	Strong	Strong	Strong	Detectable	Weak
White Fang	W	Detectable	Strong	Strong	Detectable	Detectable	Weak
White Fang	V	Detectable	Strong	Strong	Weak	Weak	Weak

Figure 140: Alternative arrangement of figure 39 from chapter four

Appendix 6 – *Loss* (2019) cast and crew

Producer/Director - David Lee
Mother - Gretchen Murray
Father - Roger Crook
Youngest Daughter - Sophie Fulton
Middle Daughter - Maddie Thomas
Oldest Daughter - Heather Patterson
Dog - Chancho

Director of Photography/Editor - Milan Marić
1st AD - Colin Biesterfield
Set PA - Russell Andrews
Set PA - Jamie Adams
Runner - Scott Gordon

Authors Contact

Email: m.maric@massey.ac.nz, markuza@gmail.com
LinkedIn: https://www.linkedin.com/in/milanmaric/