McGill, M., Williamson, J. H., and Brewster, S. (2016) Examining the role of smart TVs and VR HMDs in synchronous at-a-distance media consumption. *ACM Transactions on Computer-Human Interaction*, 23(5), pp. 1-57. (doi:10.1145/2983530)

This is the author's final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.gla.ac.uk/132610/

Deposited on: 15 December 2016

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk

Examining The Role Of Smart TVs And Virtual Reality Head-Mounted Displays In Synchronous At-A-Distance Media Consumption

Mark McGill, University of Glasgow
John H. Williamson, University of Glasgow
Stephen Brewster, University of Glasgow

This paper examines synchronous at-a-distance media consumption from two perspectives: how it can be facilitated using existing consumer displays (through TVs combined with smartphones), and imminently available consumer displays (through VR HMDs combined with RGBD sensing). Firstly, we discuss results from an initial evaluation of a synchronous shared at-a-distance smart TV system, *CastAway*. Through week-long in-home deployments with five couples, we gain formative insights into the adoption and usage of at-a-distance media consumption and how couples communicated during said consumption. We then examine how the imminent availability and potential adoption of consumer VR HMDs could affect preferences toward how synchronous at-a-distance media consumption is conducted, in a laboratory study of 12 pairs, by enhancing media immersion and supporting embodied telepresence for communication. Finally, we discuss the implications these studies have for the near-future of consumer synchronous at-a-distance media consumption. Combined, these studies begin to explore a design space regarding the varying ways in which at-a-distance media consumption can be supported and experienced (through music, TV content, augmenting existing TV content for immersion, and immersive VR content), what factors might influence usage and adoption and the implications for supporting communication and telepresence during media consumption.

CCS Concepts: •**Human-centered computing** → **Empirical studies in HCI; Empirical studies in collaborative and social computing; Collaborative and social computing devices;**

General Terms: Experimentation, Design

Additional Key Words and Phrases: At-a-distance, dislocated, media, consumption, couples, co-viewing, sync-watching, long-distance, mixed reality, Chromecast, video-mediated communication, computer-mediated communication, synchronous, virtual reality

ACM Reference Format:
Mark McGill, John H. Williamson and Stephen Brewster, 2015. Examining The Role Of Smart TVs And Virtual Reality Head-Mounted Displays In Synchronous At-A-Distance Media Consumption. *ACM Trans. Comput.-Hum. Interact.* 0, 0, Article 0 (February 2016), 60 pages.
DOI: 0000001.0000001

1. INTRODUCTION

"3... 2... 1... play!". This ritual will be familiar to many who have synchronously watched media content with their geographically separated partner, friends, or family. The attempt to synchronize over a given communications medium, timing the press of the play button so that media sources are aligned, is commonly recanted. In the media, this phenomenon has been termed "Sync-watching" [Giridharadas 2014], however we refer to it as synchronous at-a-distance media consumption. The synchronous element can vary wildly with such approaches, with buffering of streams, pauses in playback, and shifts in attention all affecting you and your partner's relative positions in a shared media stream. The at-a-distance element too varies, from cross-residential

Author's addresses: Mark McGill and Stephen Brewster, Glasgow Interactive Systems Group, School of Computing Science, University of Glasgow; John H. Williamson, Inference, Dynamics and Interaction, School of Computing Science, University of Glasgow.
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
© 2016 Copyright held by the owner/author(s). 1073-0516/2016/02-ART0 $15.00
DOI: 0000001.0000001

friends at opposite ends of a city, to partners on different continents. The net effect, however, is invariably the same: in communicating and sharing media experiences synchronously at-a-distance, we become closer to those we watch with [Macaranas et al. 2013], and we engender greater intimacy [Dansie 2012] in our relationships.

The importance of this effect becomes apparent when we consider the scale of one particular demographic: couples in long-distance relationships. In the USA alone, there are estimated to be 7 million couples in long-distance relationships, with census data from 2005 suggesting that there are approximately 3.6 million married persons who live apart "for reasons other than marital discord"[1], for example because of economic migration or education. Indeed, as many as 75% of students in the USA are likely to have taken part in a long-distance relationship during their college education [Stafford 2005]. This is a significant portion of the population for whom technology facilitating at-a-distance synchronous media experiences could strengthen their relationships. Indeed, those that are more technologically savvy already engage in such activity, for example using web-based services such as *rabb.it* or *togethertube.com*, relying on synchronized broadcast TV content when in the same country or region, or more bespoke solutions such as manually synchronizing playback of streaming media (e.g. Netflix) over Video-Mediated Communication (VMC, e.g. Skype). These behaviours have been readily and repeatedly witnessed in research, and have been seen to apply to both couples at-a-distance, as well as friends and family [Neustaedter and Greenberg 2012; Dansie 2012].

The fact that this ritual of synchronization is prevalent is testament both to the rise in on-demand TV, and to how this scenario is insufficiently supported by technology. In TV, the social element has primarily revolved around multi-screen experiences and collocated usage [McGill et al. 2015b], companion applications [Geerts et al. 2014], and social media use alongside TV programs [Schirra et al. 2014]. Socialization anonymously on the Internet (e.g. using *twitch.tv)* is often easier than watching a specific program, with a specific person, at the same time at-a-distance. Given the advances in consumer display technology, from smart TVs to Head-Mounted Displays (HMDs), and the known effect that engaging in such experiences can have on emotional well-being and togetherness, synchronous at-a-distance experiences as-yet remain poorly accounted for in consumer technology, and consequently are not fully understood.

This paper examines synchronous at-a-distance media consumption from two perspectives: how it can be facilitated using existing consumer displays (through TVs and smartphones), and imminently available consumer displays (through VR HMDs combined with RGBD sensing). Firstly, we discuss findings from an initial in-the-wild deployment of a synchronous shared at-a-distance smart TV system. Through week-long in-home deployments with five couples (living apart within commuting distance), we gain formative insights into the adoption and usage of at-a-distance media consumption, and how couples communicated during said consumption. We examine the suitability of TV and Music content for at-a-distance consumption, the necessity of tight synchronization and the impact shared experiences facilitated in this manner have on togetherness and connectedness. In order to ensure the ecological validity of these insights, we created *CastAway*, a prototype TV at-a-distance system built on-top of Google Chromecast, a $30 smart TV dongle with approximately 17 million devices sold as of 2015[2], allowing for existing smartphone applications to be used, without modification, by multiple users across multiple geographically separated TVs synchronously at-a-distance.

[1] longdistancerelationships.net/faqs.htm#How_common_are_long_distance_relationships
[2] variety.com/2015/digital/news/google-sells-17-million-1201506974/

Secondly, we explore the implications of the imminent availability, and potential consumer adoption, of Virtual Reality (VR) HMDs, regarding how synchronous at-a-distance media consumption is conducted and experienced in the near future. VR HMDs are currently undergoing a resurgence, with headsets such as the Samsung Gear VR, HTC Vive, Oculus Rift and Playstation VR all now available, or soon to be available, to the general public. These displays may supplant use of the TV for some forms of media [McGill et al. 2015b] and may even replace the TV over a significant period of time[3][4], due to their increased immersion [Cummings and Bailenson 2015], and can be expected to see consumer adoption (e.g. in consuming broadcast TV[5]) in the near future. Combined with room-wide RGBD (colour and depth) sensing (as is often found connected to existing consumer games consoles such as Kinect for XBOX, Playstation Camera for PS4), VR HMDs have the capability to render shared mixed reality social experiences (telepresence) where those you are communicating with appear to be in the same space as yourself [McGill et al. 2015a]. Consequently, in a laboratory study of 12 pairs, we explore the extent to which enhancing media immersion (through both immersive environments for consuming existing TV media and new forms of 360° immersive VR content [Disney 2015]) and supporting embodied telepresence for communication could impact user's capability to socialize, and preferences regarding shared and synchronous media consumption at-a-distance.

Finally, we discuss the implications these studies have for the near-future of consumer synchronous at-a-distance media consumption. Combined, these studies begin to explore the design space around how at-a-distance media consumption can be supported and experienced, what factors might influence usage and adoption and the implications for supporting communication and telepresence during media consumption.

2. RELATED WORK

2.1. Social TV At-A-Distance

The TV is a hub for social interaction. Watching programmes suited to discussion, such as news and sports, is commonplace, with some programme types shown to engender feelings of community within their viewership [Bernhaupt et al. 2008]. Being geographically separated from partners can impose a significant burden on relationships, curtailing the possibility of these shared experiences.

In terms of supporting socialization at-a-distance, one option is to enable asynchronous experiences, thereby negating problems regarding synchronization of content playback across multiple geographically separated parties. For example, CollaboraTV [Nathan et al. 2008] used avatars to provide a virtual audience of synchronous and asynchronous users, with 53% of participants agreeing that the social component made watching TV more engaging and enjoyable. [Ducheneaut et al. 2008] proposed audience silhouettes as a non-disruptive means of conveying the presence of other users, whilst [Vatavu 2015] built upon this work to provide real-time audience silhouettes, where their presence affected not only users level of enjoyment, but also their own posturing and gesturing. Anonymization also makes such systems suitable to shared viewing with any currently available viewers. On a broader scale, [Schirra et al. 2014] examined the motivations for live-tweeting across a season of *Downton Abbey*, finding that the sense of connectedness such experiences provided was a significant motivating factor.

[3]polygon.com/2014/4/17/5622040/oculus-rift-project-morpheus-displays-luckey
[4]recode.net/2015/09/28/epic-games-ceo-tim-sweeney-virtual-reality-is-the-future-and-we-are-100-percent-in-qa/
[5]roadtovr.com/watched-nba-game-next-vr-never-want-go-back/

2.2. Intimacy Through At-A-Distance TV

The biggest benefits of social TV, however, are to be had when there is a deeper social link between those viewing the content. Those in relationships, familial relations, and close friends are all groups for whom geographic separation can impose a significant cost in terms of togetherness and intimacy. Intimacy is a key component of relationships, and technology has shown to be capable of playing a significant role in enabling intimacy at-a-distance [Hassenzahl et al. 2012], at times in perhaps unexpected ways (e.g. YourGloves [Gooch and Watts 2012] enabling at-a-distance hand holding). Consuming TV content with others at-a-distance is one way in which technology can play a significant role in bolstering this intimacy [Williams et al. 2009]. Other forms of media, such as music, may also be able to play significant roles in terms of at-a-distance intimacy. [Neustaedter and Greenberg 2012] found that two participants watched music videos together through synchronizing the start of YouTube clips. They noted that "both participants enjoyed seeing their partners' facial reactions to the songs and videos over their Skype connections".

There have been a number of implementations of at-a-distance and social TV systems, both in research and commercially. [Harboe et al. 2008b] presented "ambient social TV" where users could see what others were watching and send lightweight messages, whilst [Weisz et al. 2007] integrated text chat with video viewing successfully. [Harboe et al. 2008a] provided an open audio link between participants' homes, finding that social TV "added value over and above watching alone", helping to "relieve boredom and provide distraction during commercial breaks and slow segments of the show" and "enhance the intensity of the experience, such as when two rooms cheered together at an event in the game". [Palviainen et al. 2013] supported presence and togetherness through voice and text based chat, gestures with avatars and a social EPG. There is also the question of how in-sync users need to be. [Geerts et al. 2011] concluded that when using speech chat at-a-distance, users noticed differences above 2 seconds, whilst using text chat delays up to 4 seconds were tolerable.

With respect to commercial implementations, Zync [Shamma and Liu 2009] integrated synchronous sharing of video content through an instant messenger program, where users employed video as an enhancement to conversations, providing a common background as attention to the conversation varied. There have also been attempts at operationalizing synchronous media consumption, for example the former XBox 360 Netflix "Party mode" [Narcisse 2011], and sites and extensions such as *netflixparty.com*, *rabb.it*, *togethertube.com*, *letsgaze.com*, *plug.dj* and *showgoers.tv* all provide varying browser-based means for synchronizing playback of various media across multiple geographically disparate users. This theme of at-a-distance consumption has been revisited repeatedly, however as-yet no smart TV platform has readily adopted or provided a solution to facilitate this behaviour generically across content providers.

2.3. Who Do We Consume Media With At-A-Distance?

In such forms where there is no anonymity provided, who we consume this media with is likely limited to our close social connections. For example, in a workshop [Hess et al. 2011] had one participant explicitly request the ability to see what video their friends were watching at the moment. This "triggered off a critical discussion because the participants only want to involve a small subset of their buddy list". This need for a strong social connection between users reflects work by [Dezfuli et al. 2011] which found that close friends and family were those people most wanted to consume such media with. This also has implications for how shared experiences are initiated, with scheduling such events proving difficult for many [Aguila 2011], necessitating the development of

routines as to when partners would be available for each other. Technology can play a part here, for example [Metcalf et al. 2008] used ambient lights to draw attention to the TV when others in the social group were watching.

2.4. How Do We Communicate At-A-Distance?

For there to be a connection between those engaging in a shared, synchronous at-a-distance TV experience, a channel of communication is necessary. For example, [Bernhaupt et al. 2008] featured interviewees that used Video-Mediated Communication (VMC) for shared viewing of soccer matches and TV-quiz shows, in order to approximate the experience of "doing something together", with socialization aided by the shared reference point of TV. Indeed, for TV at-a-distance, VMC is often purported to be the primary means for communication, due to the intimacy and privacy this medium allows [Buhler et al. 2013]. [Neustaedter and Greenberg 2012] examined how couples communicated at-a-distance, demonstrating that the presence provided by VMC was key in providing intimacy, reinforcing findings from [Aguila 2011] regarding computer-mediated communications easing loneliness and increasing feelings of closeness, and [Dainton and Aylor 2002] regarding relationship satisfaction. In interviews, seven participants watched television or videos together, using a laptop placed near to, or in front of, a couch such that they could broadcast their reactions. VMC was also used during other parallel activities, e.g. eating dinner, reading, and gaming. The importance of these shared experiences was emphasized by [Brubaker et al. 2012], with one participant describing a period of 4.5 years in which he and his partner used Skype to enact movie date nights to maintain their relationship. In a survey of 24 professionals that relied on VMC in their personal and professional lives, they found that 57% of participants had used VMC to share activities with others, including "attending parties (22%), family events (32%), and watching TV or a movie (26%)".

[Macaranas et al. 2013] examined the usage of VMC for at-a-distance video consumption in three parts. In a survey (106 respondents), approximately a quarter of respondents had tried sync-watching at least once, with another quarter expressing interest in trying it, with a bias in these responses toward younger age groups. In a field study (56 participants, intimacy pairs), they had participants schedule a time with their remote companion to watch together. 15 minutes prior to watching the program, participants were expected to log in to Skype and initiate a video chat with their partner, with synchronization achieved by starting the video playback at the same time manually. Finally, in a lab study, they examined the effect the viewing location had on the video-mediated communications experience, comparing Local (watching TV in the same room) to Picture-in-Picture (PiP, with their partner inset on the TV), and Proxy (with their partner on separate device) Conditions (see Figure 1). The found that PiP was rated the least enjoyable, with no significant differences between Local and

Fig. 1: The lab study conditions from [Macaranas et al. 2013]: (A) watching in the same room; (B) Picture-in-Picture; (C) Proxy (remote person on separate device).

Proxy, and it had the lowest Social Presence (SP) score, with Local having significantly higher SP than PiP or Proxy. They concluded that this suggested "the communication media fidelity plays a strong role in the social connection of the experience". However, the results of the lab study were contradicted by the field study where, given the option of selecting which configuration out of PiP or Proxy to use, 61% of participants opted to use the PiP configuration, with no significant differences between Proxy and PiP found in terms of enjoyment. Moreover, they found that participants experienced a high degree of connectedness in the field study, ascribed to the common ground and shared activity of the video experience.

Macaranas *et al.* suggested that the next step in such work would be to develop software and/or hardware to support watching together remotely, suggesting that there lay challenges in "initiating the experience, choosing the program to watch, closely synchronizing playback, and solving audio crosstalk". Furthermore, they suggested that "watching TV is but one of many possible remote shared experiences. This study strongly supports rich media beyond audio communication in remote shared experiences. This is a rich design space that deserves more exploration".

2.5. The Role Of Mixed Reality HMDs

Much of the research discussed was predicated on the assumption that the TV would play a central role in media consumption, alongside other similar "second screen" displays such as smartphones and tablets, with communication being facilitated, in some form, by whatever displays were available in the multi-screen home. However, the advent of consumer Mixed Reality HMDs [Milgram and Kishino 1994; Milgram and Colquhoun 1999] which blend the real and virtual, be they Virtual Reality (VR) displays with some integration of cameras and sensing to allow for Augmented Virtuality rendering (such as the Oculus Rift [Oculus 2015]), or Augmented Reality (AR) displays (such as the Microsoft Hololens [Microsoft 2015] optical see-through HMD) which render over reality, has led to the possibility of TVs being supplemented, or even supplanted, for some experiences [McGill et al. 2015b]. In the case of VR HMDs, these displays have the potential for increased immersion in a given experience through the inducement of presence in a virtual world i.e. instead of watching a film or playing a game through the window that is the TV, instead these media types can be experienced as if the viewer was actually there, to a degree determined by a variety of factors (e.g. rendering fidelity, headset fidelity, tracking etc. [Cummings and Bailenson 2015]).

VR experiences can be dynamic and 3D rendered, or static (e.g. omni-directional 360° video such as [Disney 2015]), however the effect is largely the same: an illusion of place and plausibility is created, with the viewer inhabiting an Immersive Virtual Environment (IVE) [Slater 2009]. In the case of AR HMDs, currently they exist primarily to augment and selectively occlude reality rather than supplant it. For example, instead of having a TV or display of a fixed size, in a fixed location, AR HMDs make it possible to dynamically instantiate such displays and have immersive experiences that are rooted in reality[6]. Considered from the perspective of media consumption, there are strong arguments to be made that VR and AR HMDs might supplement or supplant the TV. They allow for more immersion that is possible with existing TVs in the case of current VR HMD-based experiences. And they allow for TV experiences that can adapt to the available real-world environment (in terms of presentation size, following the user, augmenting the environment to match the experience [Campbell et al. 2014] etc.).

[6]e.g. the Microsoft Hololens demo where an AR TV follows the viewer, and changes in size based on user inputs, see www.youtube.com/watch?v=3AADEqLIALk

2.5.1. Consumer Adoption. VR HMDs have existed for decades. Yet, as Hutchison noted regarding their advent as a consumer reality "Nicholas Negroponte predicted in 1993, in Wired Magazine, that head mounted displays would be ubiquitous by 1998... Clearly, in 2007, head mounted displays are conspicuous by their absence from every-day use". However, the potential for consumer VR adoption has increased markedly over the past few years, given the recent increase in the availability of affordable, comparatively high-fidelity consumer VR HMDs across various platforms. For example, mobile VR (e.g. Google Cardboard, Samsung Gear) is now supported by any relatively recent smartphone, whilst high-end consumer VR (e.g. Oculus Rift, HTC Vive) has recently become available to PC users.

However, most pertinent to this paper are the advances in living-room VR experiences that are expected to reach consumers within the next year. Consumer gaming consoles are, in effect, high-performance networked living room PCs that have seen widespread adoption in the home (e.g. 35.9 million Sony PS4s have been sold worldwide[7]). In addition, a proportion of these consoles are equipped with RGBD cameras which provide imaging of the living-room (e.g. Sony Playstation 4 Camera[8], XBOX One Kinect[9]). In the case of both Sony and Microsoft, both manufacturers have announced VR HMD support for their respective platforms (PS4[10] and XBOX[11]), as well as hardware refreshes of their consoles to increase their rendering capability (necessary in order to render VR experiences at a high enough resolution and frame rate to provide an illusion of presence). Over the course of the next year consoles such as the Sony PS4 and Microsoft XBOX will offer a viable route toward consumer adoption of mixed-reality VR HMD experiences, with a significant installed user base already available. Such adoption has the potential to change the nature of living-room media experiences, given that it may undermine use of the TV for immersive entertainment media consumption (e.g. films, gaming). This adoption could have implications in-turn regarding how at-a-distance media consumption is facilitated.

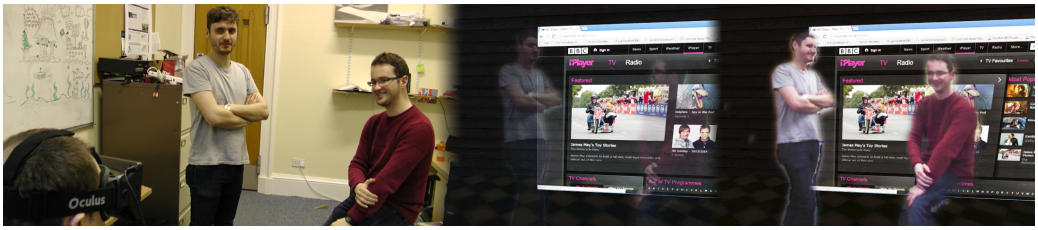
2.5.2. Social Awareness. When considered from the perspective of social use at-a-distance, the advantage these mixed reality HMDs have over the TV in terms of augmenting or rendering a world becomes more pertinent. These displays have the ability to communicate presence at-a-distance (telepresence) such that the person(s) you are corresponding with at-a-distance can be embodied in your real or virtual world in 3D, with social (such as gaze direction) and emotional (i.e. facial responses, body language) cues conveyed. This, in turn, allows for communication in a way that approximates how we communicate in person. This embodiment can take various forms, for example that of an abstract avatar, or a real-time portrayal of the at-a-distance correspondent (e.g. captured using RGBD cameras). For example, with VR HMDs, [McGill et al. 2015a] built upon prior mixed reality and telepresence research by using computer vision combined with a Microsoft Kinect to bring collocated proximate persons in reality into virtuality based on user engagement (see Figure 2), solving the problem of occlusion of others in the local environment.

In research, telepresence techniques have also been used successfully in augmented reality environments, for example [Beck et al. 2013] enabled groups at-a-distance to explore a virtual cityscape together, finding that "mutual understanding of pointing and tracing gestures independent of whether they were performed by local or remote participants" was had. More recently, [Pejsa et al. 2016] enabled life-size telepresence

[7] polygon.com/2016/1/5/10717142/ps4-lifetime-sales-35-9-million-holiday-2015
[8] playstation.com/en-us/explore/accessories/playstation-camera-ps4/
[9] developer.microsoft.com/en-us/windows/kinect/develop
[10] technologyreview.com/s/601702/why-oculus-and-htc-need-to-watch-out-for-sony-in-vr/
[11] slashgear.com/project-scorpio-official-xbox-one-with-better-everything-and-oculus-rift-13444072/

(a) Room2room AR one-to-one telepresence

(b) AR Group-to-group telepresence



(c) Mixed reality rendering of proximate persons in VR

Fig. 2: Examples of at-a-distance telepresence: (a) Room2room [Pejsa et al. 2016] enabled one-to-one telepresence through an augmented reality projection of the Kinect-captured user onto furniture, to provide some semblance of depth (b) [Beck et al. 2013] enabled group-to-group telepresence, again using projection and Kinect-based capture onto flat projection surfaces (c) [McGill et al. 2015a] conveying the presence of proximate persons to the VR HMD wearer, captured via Kinect and placed in VR based on their position in reality relative to the VR HMD wearer.

using a projection-based augmented reality space for one-to-one interactions, finding that "face-to-face communication is superior in terms of task completion, time, sense of presence, and efficiency of communication" compared to Skype-based VMC. However, for a conveyance which most closely matches reality, HMDs are currently best suited in this regard, with stereoscopic rendering allowing for a perception of presence with depth. However, this telepresence research emphasizes that having the ability to perceive an at-a-distance partner in ways that approximate how they would appear in reality can prove significantly beneficial to mutual understanding and interaction.

2.5.3. Social VR Experiences. Whilst we have the capability to convey presence through mixed reality at-a-distance, we do not yet have a full understanding of the implications of this. As [Madary and Metzinger 2016] suggested:

> "Instead of treating VR and related technologies as a replacement for in-the-flesh interaction, we should think of them as providing opportunities for new and perhaps enhanced modes of human interaction... the technology should be developed with an eye toward 'expanding and reinventing our sense of body and action'... [which] could plausibly enhance embodied (though mediated) social interaction."

Indeed it is already possible for consumers to take part in shared at-a-distance VR experiences which are mediated through shared TV and movie content (see Figure 3).

(a) CineVEO

(b) Oculus Social Alpha

(c) Convrge Cinema

Fig. 3: Examples of at-a-distance VR HMD Cinema experiences currently available: (a) CineVEO has viewers represented by human avatars, with viewing occurring in immersive locations (e.g. 60s drive-in, Haunted Valley) (b) Oculus Social Alpha has viewers represented by humanoid/animal floating heads, with viewing occurring in a small cinema setting (c) Convrge Cinema with user-generated 3D avatars in stylised virtual worlds. In all cases head movements are transmitted and 3D positional audio communications are possible between viewers.

CineVEO[12], Oculus Social Alpha[13] and Convrge Cinema[14] all allow VR HMD users to attend a multi-player virtual cinema screening, typically using simplified customisable abstract avatars due to the current lack of standardised sensors for externally capturing the VR user, and limitations regarding mobile HMD rendering performance. However this concept of at-a-distance VR mediated by TV and movie content remains

[12]mindprobelabs.com/
[13]engadget.com/2015/10/28/oculus-social-alpha-delivers-group-watching-to-virtual-reality/
[14]convrge.co/

uninvestigated, with no conception as to how these experiences might compare to TV-based at-a-distance consumption, nor any examination of the suitability of 360°, omni-directional content for synchronous consumption.

More broadly, there exist a variety of social VR experiences now available to consumers e.g. Oculus Social Trivia[15] where socialization is mediated through trivia gaming and AltspaceVR which enables social and shared interactive virtual spaces[16]. State-of-the-art experiences are nearing consumer availability, most notably in the form of the Facebook social app[17] which allows VR users to interact in a shared space with hand movements and gaze conveyed at-a-distance, with shared interactions made possible (e.g. users enacting a virtual selfie). It is thus reasonable to consider that combining the potential for immersive media experiences with the social awareness that a mixed reality rendering can convey might provide more engaging and intimate experiences at-a-distance than those facilitated through VMC.

2.6. Outcomes from Literature

There were two significant outcomes from this review. Firstly, there is a need to examine synchronous at-a-distance TV experiences in-the-wild. Prior studies have typically been lab-based, relied on direct observation in the home or utilized qualitative methods alone. Because of this, there is a gap in knowledge regarding an understanding of at-a-distance media consumption as conducted in the privacy of the home e.g. in terms of how it is organised, what types of media might prove effective at fostering togetherness (e.g. music or video content), how often couples might choose to consume media in this way, and how they might communicate during consumption.

Secondly, we wished to explore the implications surrounding VR HMDs supporting synchronous at-a-distance media consumption. These displays have a number of potential advantages compared to TVs. They allow for more immersive media (e.g. 360° video) to be consumed, and they allow for existing media to be consumed in new, potentially immersive settings (e.g. watching TV content in a virtual cinema). And they also allow for mixed reality renderings that have the capability to approximate reality. For example, social presence could be supported through incorporating others at-a-distance such that they appear to be seated beside the HMD user. However, it is unknown to what extent VR at-a-distance might be preferred to TV at-a-distance viewing. Moreover, the implications of immersion on socialization are not understood. Would highly immersive VR IVEs impede socialization, or impede the users capability to engage in at-a-distance consumption over the traditional duration of a film? In effect, the relevance and role of the TV in at-a-distance media consumption could be diminished or enhanced depending on the answer to these questions.

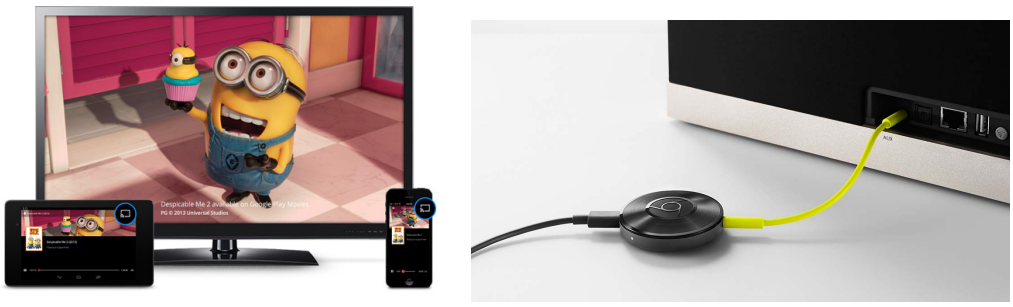
3. CASTAWAY: LOOSELY SYNCHRONOUS AT-A-DISTANCE MEDIA CONSUMPTION

The first outcome identified was to examine the adoption and usage of synchronous at-a-distance TV in-the-wild (in home), with an emphasis on ecological validity and no direct observation. In this way, we would build upon prior research which relied on laboratory studies and observed short duration field studies. To accomplish this, we developed a prototype TV at-a-distance application, called *CastAway*, which utilized a cheap and popular smart TV dongle, Google Chromecast [Google 2015].

Paired with a smartphone, a Chromecast allows for mobile applications to "cast" content to the TV screen in various forms, with support for traditional TV media (e.g. Netflix, BBC iPlayer), music (e.g. Google Play Music), games, and more. It can be con-

[15] oculus.com/en-us/blog/join-friends-in-vr-with-new-oculus-social-features/
[16] altvr.com/
[17] uploadvr.com/you-facebooks-jaw-dropping-social-vr-demo/

nected to a TV, allowing for audio-visual output, or it can be connected to speakers for audio-only output in the case of the Chromecast Audio (see Figure 4). In either case, the Chromecast streams content directly from the source via the Internet, with the role of the mobile device(s) in the room that of supplying commands to the Chromecast (e.g. what to stream / play). For example, if casting a TV program, the user would typically be presented with information about the program, and the capability to pause, seek, change the volume and subtitles on their mobile device, whilst the TV performed the playback function independent of the mobile device. A user can connect to a given Chromecast from a cast-enabled application by pressing the Chromecast icon, at which point their device will retrieve session details if the cast-enabled application is already connected, or it will close the existing application and start a cast session if the application is different (e.g. switching from casting TV content to Music). In this way, multiple smartphones or tablets can control the same session, or start new sessions.



(a) Chromecast (b) Chromecast Audio

Fig. 4: (a) Chromecast smart TV platform. Multiple devices can be connected, with the TV performing the media playback. (b) Chromecast audio dongle, which brings the same Internet streaming / casting functionality to existing speakers and audio systems.

Given the Chromecast's innate support for multi-user use, as well as its widespread adoption, low cost and availability of ecologically valid applications, we chose this smart TV dongle combined with Android smartphones as our target evaluation platform, with the aim of allowing for the same user experience as is provided in shared spaces, but at-a-distance. By this we mean that any cast commands (e.g. casting content, pausing and seeking, etc.) sent to a local Chromecast would also automatically be sent to their partner's Chromecast TV at-a-distance and *vice versa*, in effect creating a synchronous Chromecast session where the TV becomes a shared space for activity.

3.1. Implementation of "At-A-Distance" Casting

To enable synchronous at-a-distance casting and have it be transparent to existing Chromecast applications, we used rooted Android 4.4.4 phones with the Xposed framework [Rovo89 and Tungstwenty 2015]. This is a module designed to allow for system-level changes to the Android operating system, allowing applications to intercept any method call, replace or intercept returned objects, and fundamentally modify the behaviour of any application started on the device.

Using publicly available Chromecast API documentation, we set about intercepting API calls and callbacks such that we could act as a man-in-the-middle between the Android application using the Chromecast, and the Chromecast library on the mobile

device. In this way, we could act as a proxy for the Chromecast API covering parts of the `CastAPI`, `GoogleApiClient` and `MediaRouter` APIs[18] amongst others. It is important to note that this technique was used *in lieu* of having access to the source code for the Chromecast Android library; those with access to this source code (i.e. the Google Chromecast team) could much more easily and readily intercept this functionality directly, without needing the Xposed Framework.

We then built and deployed a SocketIO[19] NodeJS service hosted on an Amazon Web Services (a cloud hosting platform) instance for the Android devices to forward cast activity to, such that the activity on one device could be relayed to all connected devices, to be executed on each user's local Chromecast in order. Only then would the appropriate callbacks be made to the client cast application. In this way, we recreated the functionality of a physically shared single-Chromecast session but across multiple geographically separate Chromecasts. This change was transparent to Chromecast client applications, with the net effect of this being that user cast actions would be executed on all Chromecasts taking part in the CastAway session.

The source code for this implementation is available at [redacted]. Given the number of APIs by which applications can connect to a Chromecast, we prioritized intercepting enough of the available APIs in order for two applications to work, one for TV content (BBC iPlayer) and one for Music content (Google Play Music). Commands of relevance only to the local Chromecast (such as changing the volume, or enabling subtitles) were executed locally only.

3.2. Client Communications

We then developed a client-side application both for managing when a shared Chromecast session would be initiated with a partner at-a-distance, and for communicating with a partner in such sessions, as can be seen in Figure 5. Our application allowed for Text, Video (using the front mounted camera of the Android phone, encoded to H264) and Audio (encoded to Extra Low Delay-AAC) communications, with the functionality provided via a permanent on-screen overlay such that the chat functionality existed on top of every application. When pressing the Audio or Video buttons communication was streamed in real-time to the connected partner. Reciprocal communication was not enforced, meaning that either user could use any permutation of Text, Audio, and Video without necessitating that their partner responded on the same combination of modalities. This is unusual for VMC, given that VMC is typically reciprocal and constant (e.g. a Skype video chat), however this was chosen in order to give users flexibility and allow fine-grained instrumentation of their communications - they could communicate and respond using whichever modality they wished, and we in turn could examine in detail how couples chose to communicate. To emphasize immediate applicability, practicality and ecological validity, no additional cameras or hardware were required to use the system, with all communications mediated via the mobile devices.

A permanent notification (see Figure 6) provided users with the ability to mark themselves available to sync, and to see if their partner was available to sync, with audio and vibration notifications used when this state changed. When in the TV and Music applications, this notification changed to provide a limited set of functionality for managing the session, showing users what content was playing, how well synchronized their media stream was to their partner's, what application was connected, and allowing them to pause/play the media content (thereby pausing/playing for their partner also) and re-synchronize the streams if the delay became noticeable. Aside from this, no other interventions were provided regarding synchronization. We refer to this

[18] http://developer.android.com/reference/com/google/android/gms/cast/package-summary.html
[19] http://socket.io/

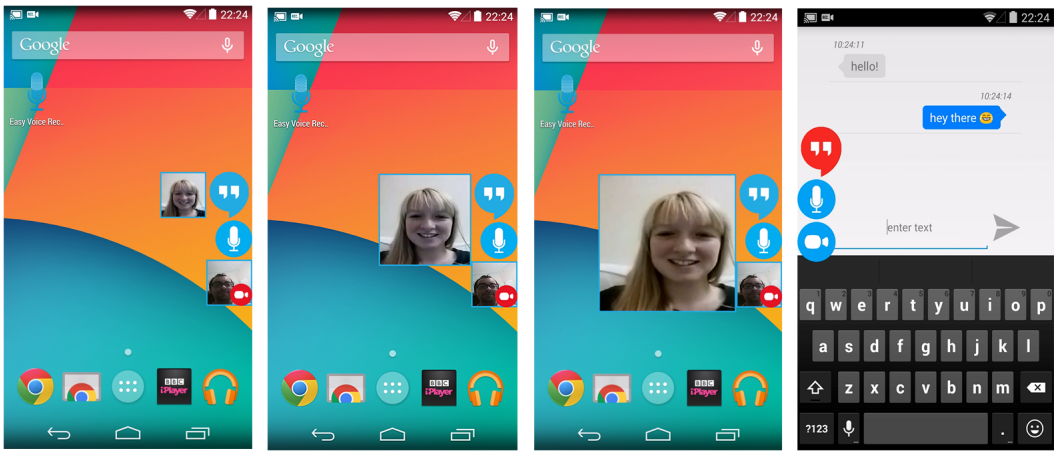


Fig. 5: The chat UI for the CastAway system. A draggable list of 3 icons allowed users to instantly chat via text, speech, or video (and any permutation of these) to their partner when connected to a CastAway session. This chat UI was permanently overlayed on the Android device, allowing for multi-tasking with on-going communication. See supplementary video for footage of the system in use.

system as being "loosely synchronous" as under healthy network conditions playback synchronization remains approximately within the guidelines laid out by [Geerts et al. 2011].

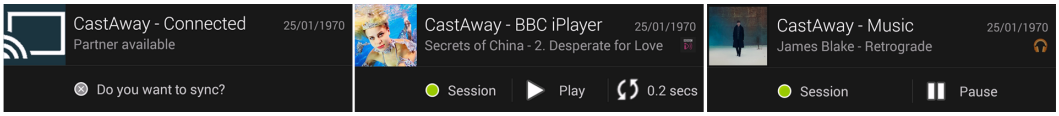


Fig. 6: A permanent notification allowed users to see if their partner wished to start a shared session, and allowed them to control ongoing sessions, for example pausing playback, or attempting to re-synchronize streams (here a 0.2 second delay is indicated between clients). Additionally metadata (art, name, application) about what content (if any) was currently being played in the synchronous Chromecast session was also displayed.

3.3. Study 1: Initial TV At-A-Distance Adoption And Usage In The Home

Our implementation of CastAway afforded a unique opportunity compared to prior research: in controlling both the communication and media functionality, we were able to instrument and measure communications and usage, in contrast to prior studies which relied on self-reported measures or video annotations. Moreover, by providing two different media experiences (Music and TV), we could examine whether different forms of media content would have the same effects and thus relevance to being consumed at-a-distance, in terms of adoption, fostering togetherness and intimacy. This allowed us to test research questions regarding how the system might be appropriated, compared to what had been detailed in the literature, specifically:

RQ 1.1. To what extent would couples choose to synchronously consume TV / Music?
RQ 1.2. What effect would synchronous TV and Music consumption have on perceived togetherness and closeness on couples, compared to communicating without a shared synchronous media experience?
RQ 1.3. How would couples choose to communicate during synchronous consumption, and would this vary depending on the media type being consumed?

We chose not to implement any form of automatic synchronization correction in CastAway. Instead, we gave users the capability to manually re-synchronize content streams through a single button press on a dialogue which informed them of the approximate time difference (in seconds) between the CastAway users when watching a TV program. This introduced one further research question:

RQ 1.4. When do users perceive the need to re-synchronize with their partner (if any)?

Maintaining client synchronization in and of itself is not problematic. However, by maintaining synchronization rigidly, an element of frustration is likely to be felt by at least one of the users. Re-synchronization necessitates that either the content being played back pauses for one viewer (whilst the other's local playback catches up), or that the content is taken back to a common time stamp (leading to one viewer re-watching content). Thus there is merit in examining if / when the frustration of re-synchronization can be avoided, by allowing an element of de-synchronization. By allowing users to self-determine when re-synchronization should occur, we can begin to examine how synchronized content playback needs to be in a real-world context, expanding upon [Geerts et al. 2011].

3.4. Demographics and Design

Couples were recruited on the basis of a number of pre-requisites: they needed to be smartphone users, familiar with VMC such as Skype, and they needed to live apart, beyond walking distance or a single public transport journey, such that they undertook, and had a need for, at-a-distance interaction. Additionally, they were to have no visual/audio impairments. For this study, 5 couples in relationships were recruited from University mailing lists (6 males, 4 females, mean age=20.9, std.dev=1.1, average distance apart=27.4 miles, std.dev=43.7 miles), all living within the same timezone. These couples all lived apart, at varying commutable distances (i.e. able to see each other 3-4 times per month), and as such represented a demographic with moderate need for at-a-distance support (as opposed to couples in long-distance relationships, where the opportunity for visitation / collocation is typically more restricted). None of the couples were married, and all couples reported that they communicated with their at-a-distance partner daily. Prior to deployment, couples were interviewed regarding how they typically stay in contact with their partner, with the majority utilizing text messaging (10 participants) and phone calls (9), followed by social networking (8), instant messaging (6) and video chat (5). Additional options for blogs, twitter, and an open "other" field received 0 responses.

This sample size was reflective of previous field studies and probes into in-the-wild behaviour (e.g. [Harboe et al. 2008b] utilized 4-5 couples/households per Condition, [Palviainen et al. 2013] utilized a field test with 11 participants, [Beck et al. 2013] examined 12 participants, [Rost et al. 2016] studied 10 participant's communications) and our aim to gather data that was more representative of real-world practice than prior studies (requiring long deployments, as opposed to shorter laboratory evaluations), by deploying CastAway into homes without direct observation or control over usage.

Each partner in the couple was provided with a phone pre-loaded with the CastAway software and a Chromecast each, and given a demonstration as to how the system operated in person if available to collect the equipment from the University. In one case where this was not possible for one participant, their partner was given a demonstration whilst they were given a detailed manual. Participants were instructed that once they had the system operational in their homes, they were to familiarise themselves with both the communications and media functionality, namely the two at-a-distance applications, BBC iPlayer and Google Play Music (hereafter TV and Music). At this point, the study would begin and the participants would have use of the system for 1 week.

3.5. Measures

All usage of the Chromecast and communications functions during the course of the week was recorded, allowing us to measure the extent to which Music and TV content was consumed, what occurred in synchronous sessions and how the couples communicated in detail. For TV content, the extent to which content was synchronized was also recorded. A "Connectedness" questionnaire was also delivered to participants at three points (meaning 30 questionnaires were filled out across the 10 participants): immediately after first usage of the communications functionality only, and again after their last usage of the TV and Music functionality. This comprised firstly of the Affective Benefits and Costs of Communication Technologies (ABCCT) questionnaire [Yarosh et al. 2014]. This is a CSCW questionnaire examining the emotional benefits and costs of a given social communications medium in field deployments, that has been shown to be sensitive enough to show differences between two communication systems and takes less than 10 minutes to complete. It has seven scales upon which a communications medium is evaluated: emotional expressiveness, engagement and playfulness, presence-in-absence, social support, feeling obligated, unmet expectations, and threat to privacy, and thus covers a range of emotional factors, and has been notably used in related work [Forghani et al. 2014]. Alongside this was the Social Presence factor from [Macaranas et al. 2013] (SP1–4, 6, and 7).

A post-study questionnaire also asked questions regarding user experience of social TV (perceived usefulness, attitude, and intention to use) from [Shin 2013], emotional connection from [Forghani et al. 2014], synchronization from [Geerts et al. 2011], and engagement / togetherness from [Inkpen et al. 2013]. Participants were interviewed after the study using questions derived from [Dansie 2012] regarding preferences and the effect that using the system had on their relationship over the week. See Appendix A for full copies of the questionnaires.

3.6. Plots and Tests

The majority of the plots in this paper are Violin plots as demonstrated by [Hintze and Nelson 1998] using the *ggplot2* library. They are plotted using *geom_violin()* displaying a rotated kernel density plot on either side of a box plot. These plots allow for density estimation, such that a more accurate understanding of the distribution of the data can be had by the reader. The thicker the plot at a particular point, the denser that particular region, whilst long thin plots describe wide distributions of data.

Each Violin plot contains within it a standard boxplot (plotted via *geom_boxplot()*). Boxplot boxes indicate the first and third quartiles (25th and 75th percentiles). They also feature notches [Krzywinski and Altman 2014] which denote the 95% confidence interval[20], essentially allowing a by-eye estimation of significant differences (when the notches of any given boxplots do not overlap). This approach is favoured by those in

[20] sites.google.com/site/davidsstatistics/home/notched-box-plots

HCI that believe all reporting should be done with confidence intervals and visualizations, as opposed to Null Hypothesis Statistical Testing [Dragicevic 2015]).

Unless otherwise stated, for parametric tests a repeated measures ANOVA was performed using *lme()* in *R* as prescribed by [Field et al. 2012], with likelihood ratios reported, and *post hoc* Tukey contrasts performed where applicable. For non-parametric tests a Friedman's ANOVA was performed using *friedman.test()* in *R*, with *post hoc* pairwise Wilcoxon Rank Sum Tests performed where applicable.

3.7. Findings From In-The-Wild Deployments

CastAway was deployed with 5 couples, each of whom had the system available for use for a full week. This availability resulted in approximately 33 hours (38 sessions) of in-home usage of CastAway, an average of 6.6 hours per couple, with usage predominantly occurring in the evenings.

3.7.1. How were sessions initiated, and what media were consumed? (RQ 1.1). On the initiation of a CastAway session, both partners were prompted with an optional question asking how the session had been arranged. Sessions were predominantly initiated on an *ad hoc* basis or scheduled on the day, as can be seen in Figure 7.

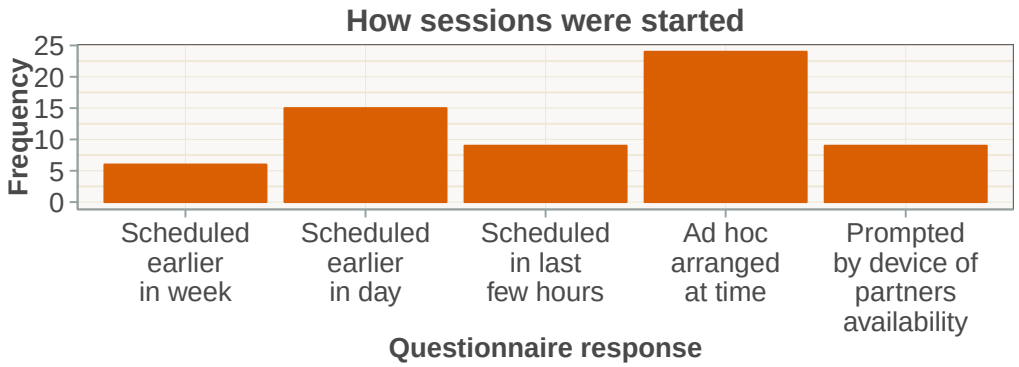


Fig. 7: How CastAway sessions were initiated across participants (optional, with question delivered at the beginning of each session for both participants).

There was one "other" response, which indicated a restart of a previous session, and one differing response between the pair, where one partner believed the session to have been scheduled earlier in the day, whilst the other partner believed it to be *ad hoc*.

A breakdown of what media was consumed over the course of the week can be seen in Table I and Figure 8, with usage shown by Comms. (communications only, being conducted at the start of the study as a baseline when not consuming media synchronously), Music, TV and Mixed (Music and TV in the same session). Sessions lasted on average for 30 minutes when consuming Music, 77 minutes when consuming TV content, and 46 minutes when mixing TV/Music.

In terms of frequency and duration of sessions, in a RM-ANOVA there was no significant difference by frequency $\chi^2(3) = 4.46, p = 0.22$, however there was a significant difference by mean duration $\chi^2(3) = 10.32, p = 0.02$, with *post hoc* Tukey's tests showing a significant difference between TV and Comms, as can be seen in Table I. Synchronous casting sessions tended to last for a minimum of approximately 30 minutes, meaning that users engaged with the system and did so on multiple occasions.

Table I: Mean total duration, mean duration of session, and mean frequency of sessions across groups, shown by: Comms (communications only, no media), Music, TV and Mixed (Music and TV in the same session). Brackets indicate standard deviations.

Usage	Mean Total Duration (h:mm)	Mean Session Duration (mins)	Mean Frequency of Sessions
Comms. only	0:07 (0:11)	4.76 (5.99)	1.0 (0.71)
Mixed	1:14 (1:40)	33.77 (34.93)	1.6 (2.51)
Music	0:49 (0:22)	33.58 (14.37)	1.6 (0.89)
TV	4:24 (3:52)	64.70 (43.17)	3.4 (2.97)

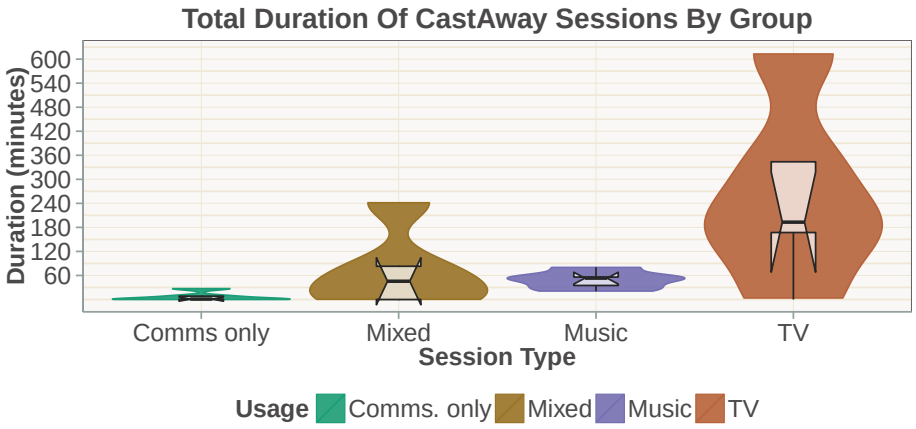


Fig. 8: Total duration of usage of CastAway by groups: Comms (communications only, no media), Music, TV and Mixed (Music and TV in the same session).

In terms of total duration of consumption, as can be seen in Figure 8, TV was the dominant media consumed. In a RM-ANOVA there was a significant difference $\chi^2(3) = 11.22, p = 0.01$ with *post hoc* Tukey's tests showing significant differences between TV and Music/Mixed/Comms. Across groups over the course of the week the mean total duration of TV was 4h:24m, greater than both Music (mean=49m) and mixed sessions (mean=1h:14m). In terms of playback of content during these sessions, each group on average viewed 7 TV programs (std.dev=7.71) with TV programs typically over half an hour in duration, meaning there were no short-form videos viewed (average TV program duration=48m:52s, std.dev=16m:53s). For Music, there were on average 19 music tracks/playlists played back (std.dev=12.05). Whilst TV content clearly dominated usage, Music saw considerable adoption, being consumed both on its own, and in mixed TV/Music sessions.

3.7.2. How synchronized were the couples? (RQ1.4). With respect to quantifying the synchronization experienced by clients, we captured both perceived synchronization, as well as logging re-synchronization events (where a user requested that playback be synchronized to a common prior time stamp via the session management notification), and capturing real-time data regarding media playback synchronization. However in the latter case, we could only do so for TV playback, and not Music, due to limitations in what was accessible from the Music application.

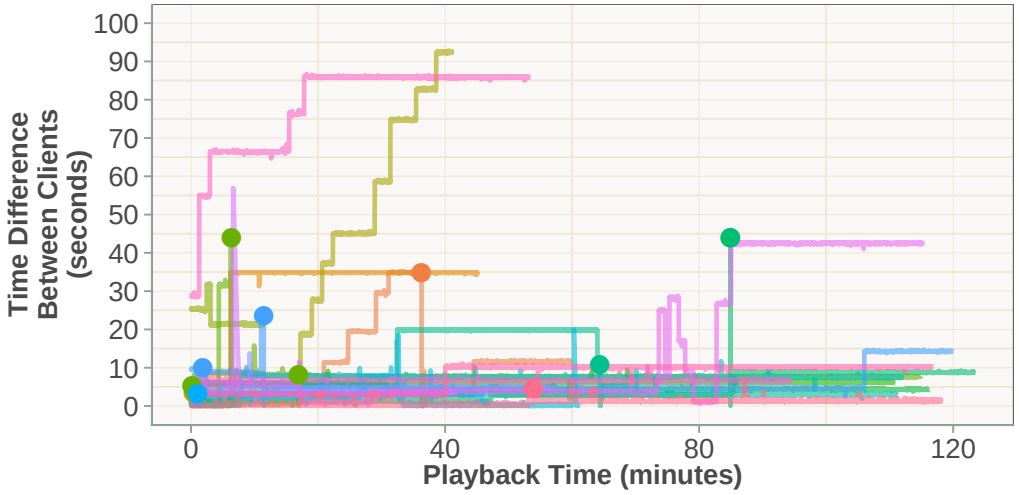


Fig. 9: Synchronization against playback time for TV content. Each colour represents a separate TV playback instance, with coloured dots indicating user-issued "re-synchronization" commands.

TV synchronization varied during the course of the study due to fluctuations in the quality of internet connectivity at the participants' households. As can be seen in Figure 9, the re-synchronization function was used by users only 12 times over the course of the study (across 26 TV playback instances), with groups using the function 2.4 times (std.dev=2.19) on average over the course of the week. This limited use is surprising considering that mean synchronization for TV playback was 11.04 seconds (std.dev=19.85), exceeding the guidelines set out by [Geerts et al. 2011]. Excluding uses of the re-synchronization function that were likely users exploring the functionality of the system (of which there were 5 occasions, at the start of playback when there was little-to-no de-synchronization) 5 of the remaining 7 uses occurred when the time difference between participants exceeded 10 seconds. There were however 7 playback instances where significant delays (ranging from 8 seconds to approximately 100 seconds) were tolerated without re-synchronization, suggesting a lack of communication-based synchronization cues (e.g. overt discussion of current events, or background conveyance of synchronization such as hearing events from a partner's stream). This is reinforced by the resultant perceived synchronization question (see Figure 10) where there was no significant difference between TV and music content, with participants strongly agreeing that their experiences appeared synchronized.

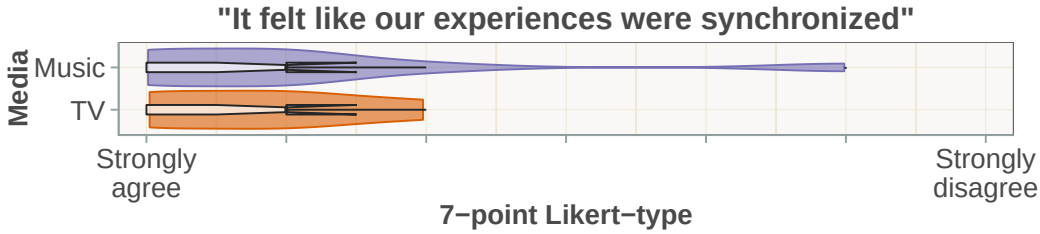


Fig. 10: Perceived synchronization across TV and Music. A Friedman's test showed no significant difference $\chi^2(1) = 0.34, p = 0.56$ between TV and Music.

3.7.3. How Did Couples Communicate During Consumption? (RQ1.3). For communication, we examined each permutation of Text, Audio and Video exclusively, meaning that, for example, Audio refers to only the audio channel being used exclusively, not including permutations such as Audio-Video. Text refers to the text chat dialogue being open on the mobile device and the device being unlocked (i.e. the time the participant was able to attend/respond to text chat); Audio refers to the participant broadcasting the phone microphone to their partner, whilst Video refers to the participant broadcasting the view of the front-mounted camera on their phone to their partner. This communication was not coupled, i.e. participants could independently choose which channels they wished to broadcast their communications over, out of Text, Audio and Video, but had no control over which channels they received from their partner (aside from leaving the session, locking their phone, or muting their phone using the volume controls).

With respect to how couples communicated, Text largely dominated across both TV and Music media types, as can be seen in Figure 11. Normalizing the total duration of communications by the total duration of usage of the system, we can see that the amount of communication per minute of usage was largely the same between TV and Music, with Text chat constituting approximately 20 seconds out of every minute of usage of the system. In a two-way RM-ANOVA there was a significant main effect on communications channel $F(6, 108) = 21.525, p < 0.01$ with *post hoc* Tukey's tests showing significant differences between Text only and Audio, Video, Text+Audio, Text+Video, Audio+Video, and Text+Audio+Video. Whilst Text chat occurred more often in Music, this difference was not statistically significant (media type factor: $F(1, 9) = 0.35, p = 0.57$).

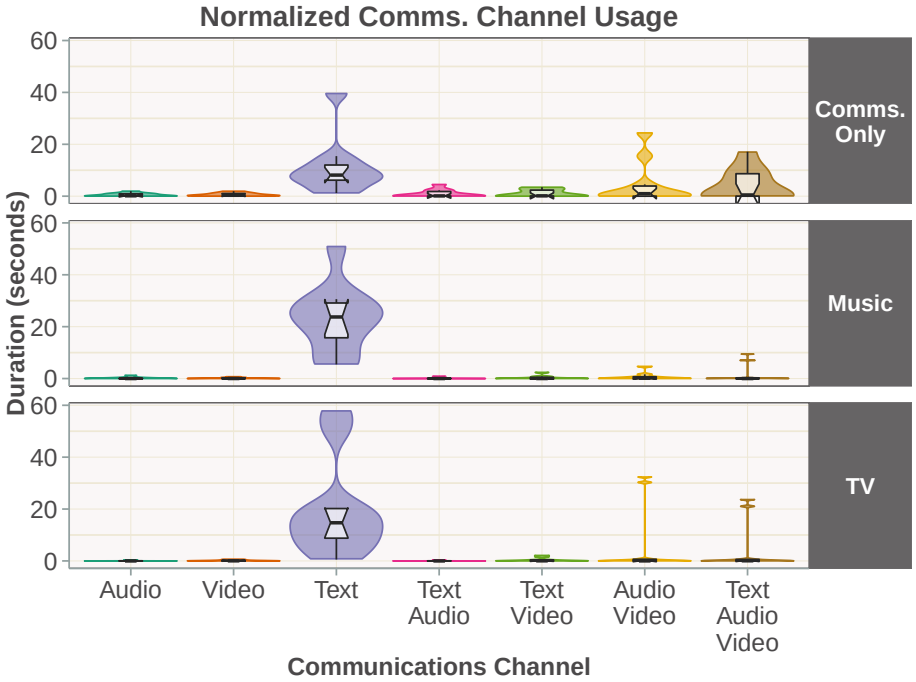


Fig. 11: Normalized usage of communications functionality per minute of system usage, $y = Total\ Usage\ of\ Comms.\ Channel\ (in\ seconds) \div Total\ Consumption\ of\ Media\ Type\ (in\ minutes)$.

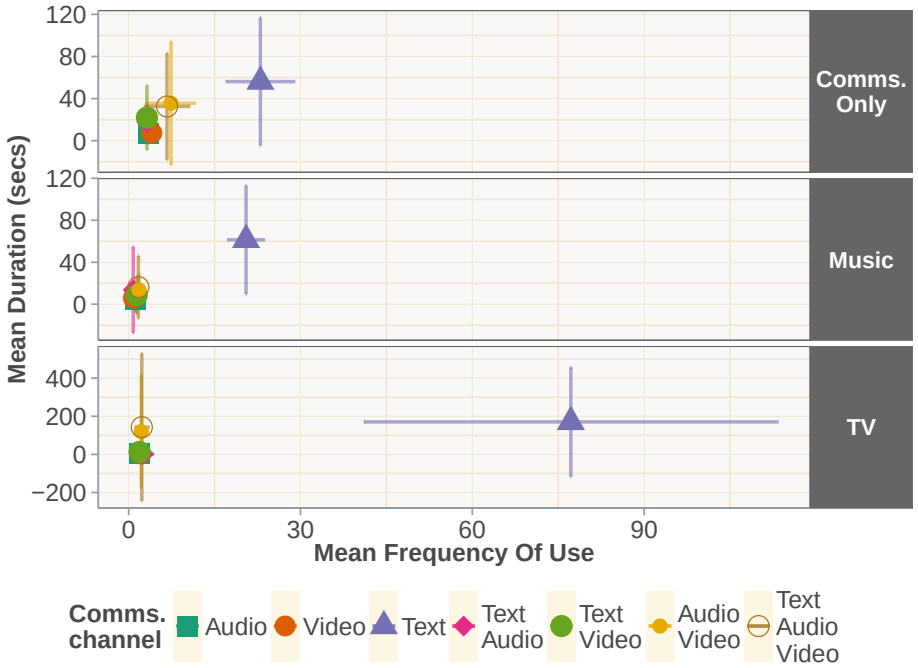


Fig. 12: Mean duration and frequency of occurrence of communications instances across groups by media type. Error bars show standard deviations.

For frequency of usage (see Figure 12), a two-way RM-ANOVA was again performed treating communications channel and media type as factors. This showed a significant main effect on communications channel $F(6, 108) = 6.74, p < 0.01$ with *post hoc* Tukey's tests showing significant differences between Text only and Audio, Video, Text+Audio, Text+Video, Audio+Video, and Text+Audio+Video, and an interaction effect $F(6, 108) = 2.37, p < 0.05$. For duration of usage there were no significant main or interaction effects. The duration of a given communication instance remained largely the same, with communication channels remaining open for relatively short periods (under a minute) when consuming Music and for longer periods (under 3 minutes) when consuming TV, with communication channels opened and closed in bursts.

In four of the five groups, Text chat was the predominant form of communication utilized, regardless of the media being consumed. The exception to this behaviour was Group A, where in 3 of their 4 CastAway sessions they relied upon Audio-Video / Text-Audio-Video for communications. For the Text-dominated groups, Text dialogs typically did not remain open throughout consumption, instead being opened and closed as required, indicative of multi-tasking behaviour on the devices. Text messaging occurred throughout consumption in all groups.

For the Text-dominated groups there were, however, short intervals when users escalated or augmented text conversations using Audio or Video communication. There were 12 occasions where participants in groups B–E utilized VMC for short intervals (4 Text-Audio, 2 Text-Video, 2 Audio-Video and 4 Text-Audio-Video). Of these, 5 uses occurred prior to/at the start of content playback, 4 uses occurred at the end of the content playback, and 3 occurred during consumption. An example of this behaviour can be seen in Figure 13.

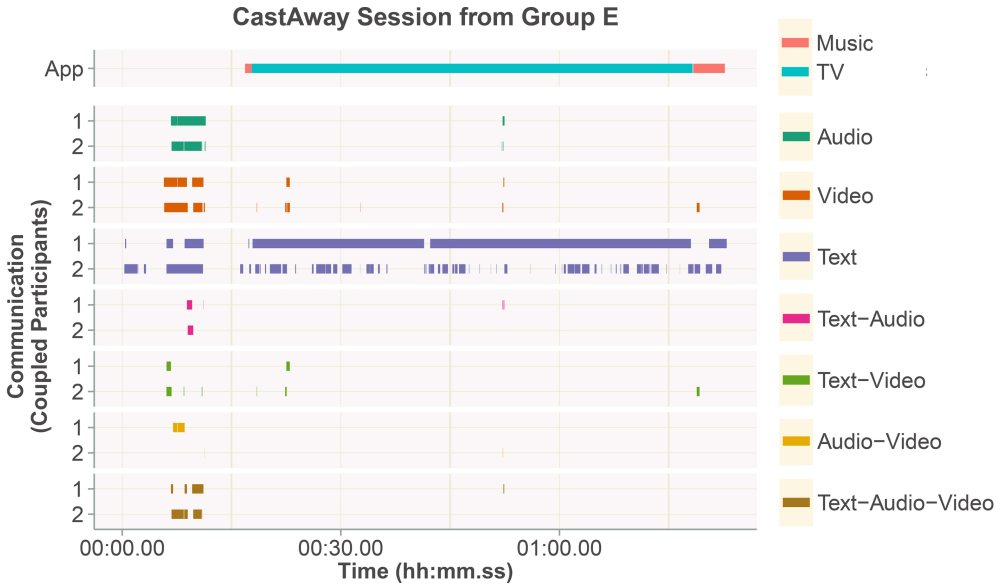


Fig. 13: CastAway session from Group E. The session lasts approximately 90 minutes, with both TV and Music consumed. Prior to the TV consumption, there is approximately 5 minutes of communication by both participants across the available Text/Audio/Video modalities, with shorter Text–Audio and Text–Video events also occurring during consumption.

3.7.4. Effect of TV and Music Consumption Compared To Communications Only (RQ1.2).

The ABCCT questionnaire [Yarosh et al. 2014] was used to compare Communications only against TV and Music usage, in order to understand if there was an effect on the benefits and costs of communication when said communication was augmented with a shared media experience. There was a significant effect on the Emotional Expressiveness subscale, however there was no significant difference in *post hoc* tests. This suggests that consuming Music whilst communicating may inhibit expressiveness to some degree, but does not confirm that this is the case. There were no significant differences on any of the other subscales (see Figure 14), with communication during TV and Music consumption seen as broadly comparable to communicating without a synchronous shared experience. This suggests that the ABCCT responses were primarily a result of the underlying communication modalities being used. Moreover, ABCCT factors appear sensitive to evaluating differences over a longer period than was evaluated in this study. Many of these factors examine infrequently occurring emergent behaviours over the course of significant periods of time. For example, our week long deployments were unlikely to encapsulate periods where social support might be needed or provided, unmet expectations might arise (given our participants were discovering how they wished integrate this form of communication into their relationships) or provide sufficient opportunity for threats to privacy as usage over a longer period of time (e.g. weeks/months).

With respect to Social Presence (”a sense of co-presence with a mediated person and an awareness for their psychological, emotional and intentional state” [Macaranas et al. 2013]) there was no significant difference across media types, again suggesting that presence was a function of the underlying communication modalities. However, for Closeness (“[Using the devices to communicate / Listening to music together / Watching

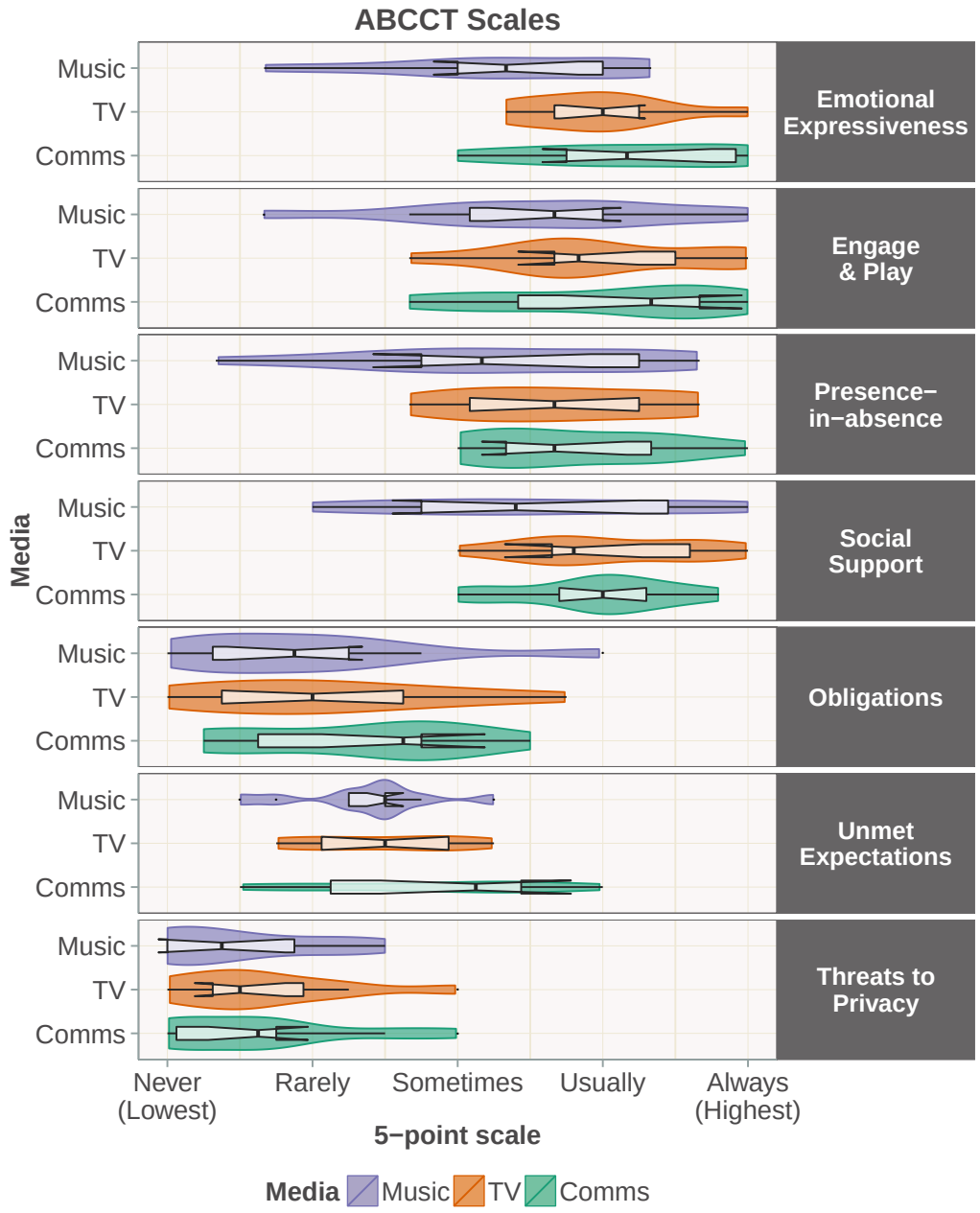


Fig. 14: Affective Benefits and Costs of Communication Technologies (ABCCT) [Yarosh et al. 2014]. Friedman's test results – **Emotional Expressiveness**: $\chi^2(2) = 6.7, p < 0.05$, *post hoc* Wilcox: No significant differences. **Engage and Play**: $\chi^2(2) = 0.19, p = 0.9$. **Presence-in-absence**: $\chi^2(2) = 0.21, p = 0.9$. **Social Support**: $\chi^2(2) = 0.38, p = 0.8$. **Obligations**: $\chi^2(2) = 3.3, p = 0.2$. **Unmet Expectations**: $\chi^2(2) = 3.3, p = 0.2$. **Threats to Privacy**: $\chi^2(2) = 1.7, p = 0.4$.

TV together] made me feel closer to my remote companion") consuming TV together was significantly perceived as helping participants feel closer to their partner compared to Communications only (see Figure 15). These results confirm that in having a shared TV experience occurring alongside communication, users feel closer to those they are communicating with. The long tail exhibited by Music is indicative of the dichotomous preference for Music at-a-distance in this study.

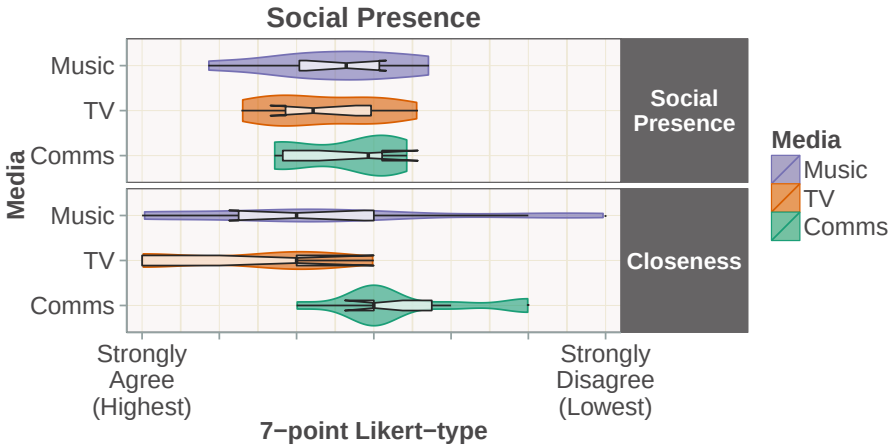


Fig. 15: Social Presence and Closeness from [Macaranas et al. 2013]. Friedman's test results – **Social Presence**: $\chi^2(2) = 2.8, p = 0.2$. **Closeness**: $\chi^2(2) = 10, p < 0.01$, *post hoc* Wilcox: Significant difference between *TV* and *Communications Only*.

Examining the CastAway system more generally, participants were asked to rate the extent to which using the CastAway system affected the couples' communications over the course of the week on a 7-point Likert-type scale (ranging from 1 to 7, 1 being "Much more / better", 4 being no change and 7 being "Much less / worse"). Whilst consuming media together synchronously did not have an effect on the perception of a partner over a communication medium, it was shown to have an effect on the couple's perceived frequency of communications (mean=2.9, std.dev=0.99), duration of communications (mean=3.0, std.dev=1.05) and quality of communications (mean=3.3, std.dev=0.67). Whilst this perceived improvement was modest, this underscores that at-a-distance media consumption can have a meaningful effect on relationships by encouraging communication and providing a shared point of discussion.

3.8. End of Study Completion Questionnaire Contrasting Music and TV (RQ 1.2)

Couples were asked to complete questionnaires examining togetherness / connectedness and their experience of social TV (e.g. in terms of intention to use in the future, usefulness, etc), for both media types (TV and Music). This was done in order to establish couples perceived differences between TV and Music when consumed at-a-distance with respect to usage and experience.

As can be seen in Figure 16, there were no significant differences between Music and TV, and their mean ratings were better than neutral, suggesting they both improve togetherness, increase the perception of experiencing activity with the at-a-distance partner, and increase engagement. However, Music typically exhibits a long tail, indicative that the perceived effect of Music consumption was not universal across couples. This trend continues in Figure 17. Again, there were no significant differences between TV and Music, with both having mean ratings better than neutral across scales. However, the long tail of Music is still prevalent.

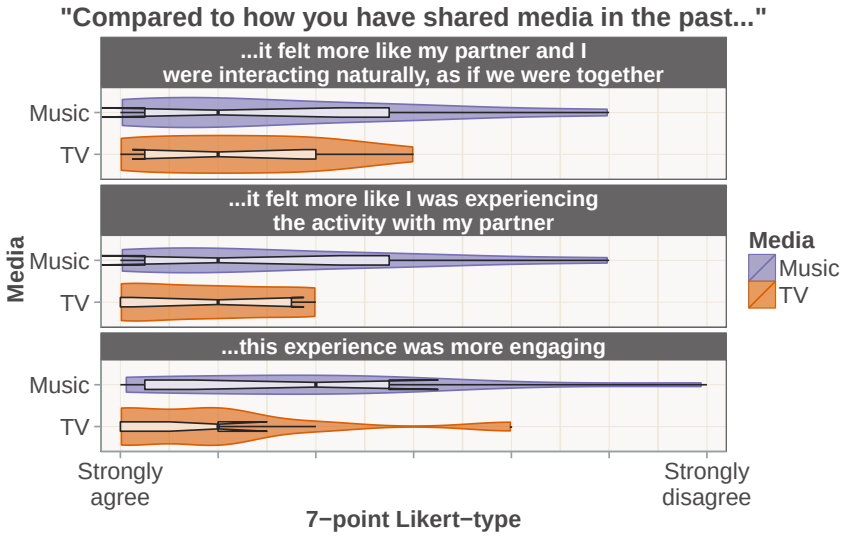


Fig. 16: Responses to "Compared to how you have shared media in the past..." [Inkpen et al. 2013]. Friedman's test results – **As if we were together**: $\chi^2(1) = 0, p = 1.0$. **Experiencing activity**: $\chi^2(1) = 0.67, p = 0.4$. **Engagement**: $\chi^2(1) = 0.67, p = 0.4$.

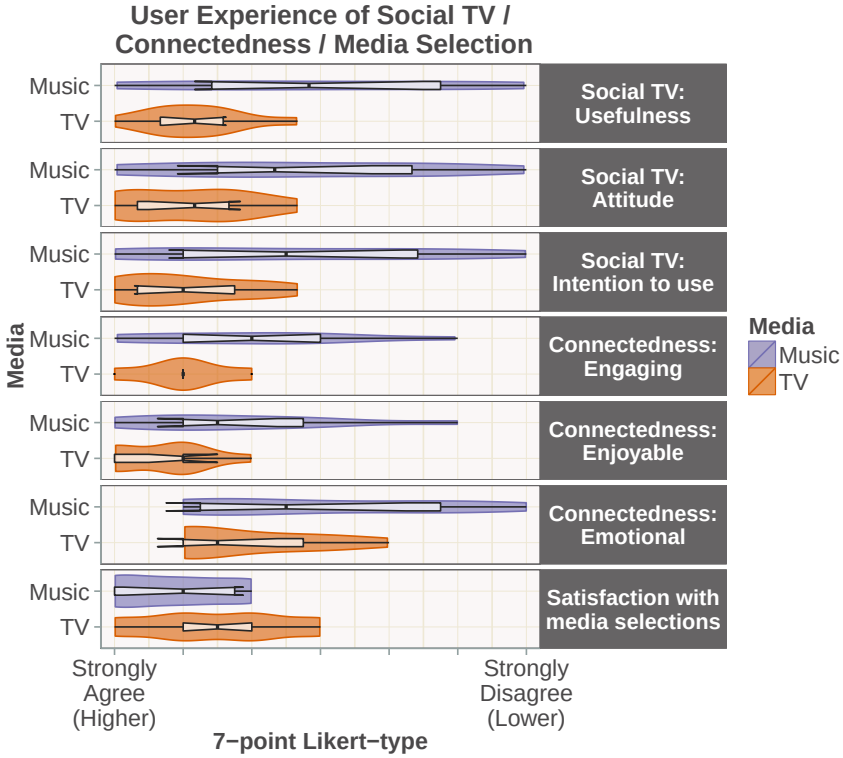


Fig. 17: User Experience of Social TV [Shin 2013] / Connectedness [Geerts et al. 2011]. Friedman's test results – **Usefulness**: $\chi^2(1) = 2.8, p = 0.1$. **Attitude**: $\chi^2(1) = 2.0, p = 0.2$. **Intention to use**: $\chi^2(1) = 1.0, p = 0.3$. **Engaging**: $\chi^2(1) = 1.3, p = 0.3$. **Enjoyable**: $\chi^2(1) = 2.7, p = 0.1$. **Emotional**: $\chi^2(1) = 4.0, p = 0.05$. **Satisfaction with media selections**: $\chi^2(1) = 1.3, p = 0.3$.

3.9. Post-Study Interview

At the end of the study, couples were interviewed regarding their experiences using CastAway. Couples were interviewed together, as opposed to individually, in order to aid recall and elaboration on points made. However, doing so incurs a potential cost e.g. in terms of bias regarding responses (e.g. mimicing a partner's feedback), the dynamics of the relationship dictating who does (or does not) respond and the extent to which a partner would be able to truthfully acknowledge issues in the presence of their partner (e.g. not wishing to use the system). These problems were partially mitigated against by ensuring that each partner had sufficient opportunity to contribute to each interview question through use of prompts and silence.

Interviews were loosely guided, with a core set of questions examining what effect usage had on their relationships, preferences regarding TV / Music use and media selections and whether the system was a burden or perceived as being intrusive. Interview transcripts were first coded using Initial Coding, where participants' statements were assigned emergent codes over repeated cycles. These codes were then grouped using a thematic approach and analysed based on the frequency and importance of the codes (see [Saldaña 2015]).

This latter step identified eight themes: *Attitudes Toward CastAway*, *Attitudes Toward TV* and *Attitudes Toward Music* related to RQ 1.1; *Role of TV*, *Role of Music* and *Effects of Usage* related to RQ 1.2; *Attitudes Toward Communication Modality* related to RQ 1.3 and *Content Synchronization* related to RQ 1.4. For each of these themes, representative excerpts are quoted, with full excerpts of all related responses provided in Appendix D. *I:* refers to the interviewer with *P#* referring to a given participant by number, *P1–2* belonging to Group A, *P9–10* belonging to Group E.

3.9.1. Attitudes Toward CastAway: At-a-distance media consumption was universally liked, but can exacerbate feelings of separation. The system proved popular with all the couples interviewed, with all five couples enthusiastic in their discussion of CastAway, expressing a strong interest in using it further and enquiring as to when the system might be released for public use. *P9–10* specifically noted that they had attempted a common bespoke solution to synchronous at-a-distance media consumption previously (using Skype and Netflix) but had run into technical problems regarding synchronization and performance, problems that were alleviated by the use of a separate smart TV system for handling the streaming and playback of content:

> *P10*: We would use it. Last year we were both on years abroad, and we tried sometimes to do that, but have [the] Skype app [connected] and watch something at the same time. But we'd always have problems, like someone's internet wouldn't work, or it would be out of sync and... it was more hassle than it's worth. But with something like this it would have been really good.
> *P9*: And also our laptops would be really slow, because you'd be running the TV program or whatever and also have Skype in the background, and it wouldn't really work. The fact this was all in one kind of made it a lot easier.

However for *P5–6*, use of the system exacerbated feelings of separation. The act of watching TV together was implicitly associated with being together, and thus breaking this association (watching together whilst not physically being together) was at times difficult to deal with:

P5: It kind of made you more aware of the fact that you weren't just sitting watching TV together [in person]. But it was nice, it felt that you were making time for each other, sitting down and watching something together.
I: Did it make it worse at all that you were being reminded that you weren't in the same space?
P6: At times yeh... When we were watching that we're really enjoying, and having proper conversations about it, it kind of reminded you that it'd be nice if you were actually together.
P5: But I suppose you don't have any other option. It's not as good as actually sitting together watching something, but it's better than watching stuff at different times and talking about it later.

This aligns with the observation by Neustaedter and Greenberg [Neustaedter and Greenberg 2012] regarding intimate activities, namely that the couple "realized more fully that they couldn't actually touch each other and this caused them to miss each other more". However in this case the couple acknowledged that the more intimate contact made possible by CastAway was still preferable. This reinforces that enabling intimacy at-a-distance may prove to be more frustrating for a subset of users, given the limitations of technology, and may require a period of adaptation when transitioning to at-a-distance experiences from in-person interactions.

3.9.2. Effects of Usage: Increased connectedness and communications. Couples agreed that the system made them feel closer and more connected to their partners, and that the system was not a burden or perceived as being intrusive during the week long deployments. CastAway lowered the cost of establishing a shared, synchronous media experience to the extent that our participants rapidly adopted it into their lives over the course of the week-long deployment, and expressed a wish to continue using it at the end of the study. It augmented their existing communications by providing a shared and synchronous reference point (TV content) or a shared backdrop to other activities (Music):

P5: It was good, it was nice, the idea that you'd make time for each other, we'd watch a lot of the same stuff anyway, it was nice to actually sit down and watch it together and talk about it as you were watching it.
P6: It helped... We have a lot of communication anyway, so it's not like it added more communication to our relationship, it was more the point that we were actually doing it, watching something, together.

3.9.3. Attitudes Toward TV: TV at-a-distance was universally liked. This encapsulated codes regarding *preference for TV*, *preference for platform and content*, and *media selection*. TV content specifically proved universally popular with the couples. The synchronization and communication functions brought them closer together and allowed them to share an activity that would otherwise have either been conducted apart and then discussed at a later date, or saved for a future meeting (*P3–4*). Moreover, the synchronization of content provided context for conversations:

P5: We talk about TV, but we'd talk about it after as we'd watch it at different times.
P6: But it's easier to sit and talk about it if you're watching it at the same time, like if he was to laugh at something I'd know exactly what he's laughing at, rather than a whole paragraph to explain, because we were watching at different times.

However two couples indicated that the limitation of BBC iPlayer was frustrating, given their personal usage of other platforms such as Netflix:

P5: It'd be good if we could have had Netflix or something as well obviously.

P8: Only if it was more than BBC iPlayer, because we ended up having to watch things we wouldn't normally watch

The complaint regarding BBC iPlayer being insufficient is an important point to note. Whilst Netflix or the BBC could implement the functionality of CastAway within their own applications, this would not be a satisfactory situation for users (limiting what content they can experience together) nor the content providers (having to re-implement functionality to match competitors and user expectations). This emphasizes that a generic solution, agnostic of content provider, is required if at-a-distance media consumption is to have the best chance of significant adoption. This justifies the design of CastAway, which was developed to be content provider agnostic and enable at-a-distance functionality transparently with respect to content provider applications. Media selections were discussed immediately prior to viewing for two couples, with one couple (Group A) having selections made explicitly by the lead partner due to a conflict in media preferences, and two other couples (Groups B and C) taking turns.

3.9.4. Attitudes Toward Music: Adoption of music at-a-distance was dichotomous. This encapsulated codes regarding positive and negative attitudes to the adoption of Music at-a-distance. Two couples were positive about consuming Music at-a-distance, and relied upon the Music functionality heavily. Of these couples, one featured music tastes that were similar, in contrast to their tastes in TV content, thus predisposing them to prefer shared Music experiences. For the other couple, a knowledge of each other's likes and dislikes coupled with the relatively low attentional engagement of Music for that couple allowed them to play Music for each other as a gift giving exercise, reinforcing connectedness through a shared audio space:

P7: That [Music] was good, because normally if we're sitting studying in our own houses, we're not really connected. But allowing us to listen to the same music at the same time helped make us feel closer.
P8: And speaking about the same kind of music.
I: Did your tastes in music differ?
P8: He put Westlife on for me for a wee bit!
P7: There's music I like and music you like, and there's music we both like, so we just stuck that on.
P8: We know what bands we both like.

In contrast, for the remaining three couples opinions on Music were mixed or negative, each for a different reason. For Group B, Music was a medium that needed to be consumed in person. For one partner specifically (P3) Music demanded a higher level of engagement, which was not suited to communicating via the provided phones. For Group C, Music tastes differed to the extent that there was a lack of common ground between the pair when it came to selecting what to listen to, in contrast to TV where their preferences were more aligned. For Group E, Music was personal, with participants preferring to experience it individually, in contrast with TV which was perceived as being more socially oriented and appropriate for discussion:

P10: I didn't like the music as much, but I feel like we probably wouldn't use it.
P9: We don't really listen to music together that often.
P10: I guess it depends on the person. Like I think music is quite a personal thing.
I: Would you say you have similar or divergent tastes in music?
P9: Similar. I don't think it's to do with the tastes, I think it's just that I

don't listen to music with other people anyway, it's a thing I do to relax.
P10: And then I guess that's not necessarily at the same time that I would want to listen.
P9: Television or films made more sense. It's more of a community based thing.

In effect, Music appears more highly personal, with less common ground, and lacks capability of TV content to drive discussion. However, where tastes are aligned, or differences tolerated, Music offered a low-engagement means of connecting with a partner. Where this is not the case, Music became irrelevant as a shared medium. This suggests Music is dichotomous in terms of adoption for at-a-distance sharing, lacking the broader appeal of TV content.

3.9.5. The Role of Music: A low-attention, mobile background activity. However, for those whom Music is suited to co-consumption, it fulfills a very different function from TV content, providing connectedness without the same attentional demands as TV content. Indeed, for the two couples that were positive about Music, it was used both as a background activity to a secondary task, and as a means of implicitly (through one partner selecting tracks for the other) forming a connection whilst requiring less attention or engagement than TV. Music did not demand seated attention, and thus allowed a freedom to move and focus on other activities:

P8: You could walk away from it and come back in and talk about it again. Music is more... walking about.
P7: Yeh, you could just have it on in the background almost.
I: Is that what you used the music for?
P7: Music was more when we were doing stuff.
P8: I was tidying my room and stuff, and I could just hear the music playing through the TV.
P7: I think that's the nice thing, that's why it's good. With the music it brings you closer.

3.9.6. The Role of TV: A sit-down activity demanding attention. Conversely, TV was a sit-down activity for all our couples, demanding more attention. This in part is likely attributed to the conversational engagement with the media, with TV content having a narrative, and on-going events which can provoke discussion in and of themselves. In order to be able to competently take part in this discussion, some measure of attention must be paid to the TV activity. As a consequence, for one couple, this led to a case of "Butler lies" [Hancock et al. 2009], text-based deception:

P2: We sat down to watch it
P1: We both sat down to watch it.
I: So you didn't have it just playing in the background?
P1: No, we just sat down to watch it, and spoke to each other while it was on. Depends what it is though, if it's something I'm interested in, I will watch it, but if it's something I'm not interested in, I'll tell him I'm watching it, but I won't watch it! (laughter)

This emphasizes that, whilst there is in essence a contract between participants that they both taking part in a synchronized media experience, implicit measures of gauging the attention being paid to said experience (e.g. gaze, body language) are fundamentally denied to users in a system such as CastAway, with explicit measures (e.g. based on conversational cues) open to misinterpretation or deception.

3.9.7. Attitudes Toward Communication. This encapsulated codes regarding comments on text, audio and video communications. With respect to communication, text communication was preferred by four of the five couples, for varying reasons. Familiarity with text communications, the lower level of engagement text requires, social acceptability, the suitability of audio communications during TV consumption, and the size of the video view all appeared to impact this preference, with reasons varying between couples:

> *P3*: Yeh, the [Text] messaging was the one we used the most, just because the video was quite small and again if you're watching something, you don't want to have a video up of someone else most of the time. And if you were going to video someone you'd use a different application to do that.
>
> *P5*: We used Text, we tried them all, but we didn't use any except text, when you're trying to watch something [it was better]... We don't really chat on the phone that much, because we're both quite busy. Especially if you're watching something.
>
> *P7*: It was easier, we were more constantly texting, but we were sometimes talking over video depending on what we were doing.
>
> *P9*: [We used] Text. I don't think we used audio at all... when you're watching a TV program you've already got a visual element and a vocal element, so the text was perfect.

However, whilst it was noted by one participant that the size of the video communications on the smartphone may have impacted usage, this preference toward text communication over permutations of text, audio and video communication appeared to often be an explicit choice regarding the suitability of the communication medium being used whilst consuming TV and Music, rather than a deficiency in the way video/audio communication was enabled. Four groups commented on how our implementation of video chat was preferable to existing implementations, specifically allowing for conversations to quickly escalate to higher fidelity VMC as and when necessary, then return to a lower-engagement means of communicating:

> *P2*: They were quite good, because you could make [the video] as big or small as you want... it was so much quicker to just connect everything.
> *P1*: For just talking to each other, it was a lot easier than using Skype or any of those things.
> *P1*: And you could have the audio, or the video, I could leave the phone on my bed and walk about the room while he's still talking to me, and do other things.

3.9.8. Content Synchronization: Synchronized content and control allow implicit communication of availability. Regarding media synchronization, there were no negative comments. Surprisingly, given the measured variability in synchronization, couples were satisfied with the perceived synchronization and the level of control over re-synchronization. Notably, the ability to unilaterally pause the experience was considered important, as it allowed participants to simultaneously avoid missing part of the content, and implicitly signal their temporary unavailability to their partner:

> *P5*: Sometimes we'd be watching something for about an hour, and it was like 9 seconds out roughly, but we didn't really bother [re-synchronizing] because it was close enough that you were watching something. Whereas if you were watching something normally, and one person pauses it for a break then you'd be 10 minutes out! It was good we could both pause it.
>
> *P7*: Almost knowing it was synchronous is better.
> *P8*: And pausing it when we walk away, like sometimes I pause stuff and leave the room and then he's in front of me and I'm behind, it was better that we were together.

This satisfaction may have been undermined had the participants preferred to use a communications medium which more accurately and quickly conveyed differences and discrepancies in playback. However, the reliance on textual communication appeared to largely insulate our participants from perceiving the time differences between their playback instances.

4. DISCUSSION OF STUDY 1

This study provided insights into the initial adoption and usage of synchronous TV and Music at-a-distance over the course of week-long deployments in the home. We instrumented usage and communication, delivered questionnaires *in-situ* examining the benefits and costs of communication technologies, perceived social presence, user experience of social TV, emotional connection, synchronization and engagement / togetherness, and interviewed participants about their experience using CastAway. With respect to interpreting these results, we discuss the limitations of this formative study, and the applicability of these findings with respect to at-a-distance media consumption within the wider population. We then go on to discuss the implications these findings have regarding our research questions, and more broadly regarding the design of TV-based at-a-distance media consumption technology.

4.1. Limitations

It is important to note that interpretation of the findings of this study should be tempered, as there were a number of technological and experimental design limitations which will have affected both the results, and the external validity of said results.

4.1.1. Design of Communications Functionality. Regarding the design of the communications functionality, this study examined a specific smart TV configuration, in the form of Chromecasts coupled with smartphones, as this represented a highly-prevalent ecologically valid configuration that could be deployed in-the-wild with minimal infrastructure or setup. As such, communication was facilitated through the supplied smartphone only, using the front-mounted video camera, in-built microphone and speaker. Communication through larger devices (such as tablets), separate anchored displays, or picture-in-picture was not considered. Nor was VMC considered where additional capture equipment was used e.g. capturing video from an alternate viewpoint in the room, capturing audio using environmental microphones. Thus results regarding communications functionality must be interpreted with these restrictions in mind, given the possibility that altering any one of these factors may have resulted in different communications usage during the study.

4.1.2. External Validity. Regarding external validity, whilst this formative study captured 33 hours of usage in total, in real-world settings, without direct observation and with high ecological validity, adoption and usage was evaluated with only 5 couples, all within a similar age group, and all within distances of each other that were within commutable distances (e.g. public transport links). This is in contrast to prior work

such as conducted by Macaranas *et al.* which examined 10 couples over short durations under observation. Given the resources available, this study prioritized fewer couples, examined for longer periods of time in-the-wild. Thus, statistical tests indicating significant differences do so only for so far as indicating differences within this sample group, and cannot be seen to be having significant external validity regarding the generalizing to the wider population and varying demographics (e.g. age, gender, relationship type, duration, media preferences in couples, cultural background, existing preferences regarding communication, potential for meetings in person, distance separated, duration of separation etc). Instead, this study should be seen as providing initial insights into the adoption, usage and design of synchronous TV and Music at-a-distance for a narrow subset of potential users, in order to guide further research in this domain.

4.1.3. Duration of Deployments. Regarding the duration of deployments, with deployments lasting approximately 1 week per couple, there was likely to be a period of novelty regarding usage and adoption. Given this, the findings from this study would be best characterised as being based on initial adoption of synchronous at-a-distance TV consumption, prior to longer term adoption and usage (which, over longitudinal deployments, could vary).

4.2. Implications for Research Questions

RQ 1.1. To what extent would couples choose to synchronously consume TV / Music?

Firstly, with respect to consuming TV at-a-distance, both quantitative and qualitative findings suggested a significant adoption of TV at-a-distance, with couples consuming approximately 4.5 hours of TV content on average over the course of the week and indicating strong preferences toward synchronous TV content at-a-distance. In contrast, Music was dichotomous. For 3 couples, Music either did not interest them (where preferences regarding Music selections clashed) or was an inappropriate form of media to be consumed at-a-distance (e.g. due to a lack of discussion content compared to TV). However, for the two couples where synchronous Music at-a-distance was regularly utilized, it was used to fulfil a different purpose from TV content, instead providing a lower-attention background activity whilst allowing other activities to be performed. This suggests that, whilst TV content has broad appeal for this use case, there may be a niche for at-a-distance activities that require less attentional demand, and better fit alongside existing activities.

RQ 1.2. What effect would synchronous TV and Music consumption have on perceived togetherness and closeness on couples, compared to communicating without a shared synchronous media experience?

Secondly, having a synchronous media experience as a backdrop to communications led to greater perceived togetherness, closeness (for TV only compared to the baseline), frequency and quality of communications. However there was no significant difference found for the costs and benefits of communication technology, nor social presence. In effect, consuming media did not change the perception of the underlying communication modalities, but did improve the prevalence of said communication. Moreover, TV content was perceived as fostering greater emotional connectedness compared to Music content, reflective of both the lack of adoption of Music content across three of the couples, but also potentially the nature of the content being consumed (with Music content often shorter and featuring less capacity for narrative engagement).

RQ 1.3. How would couples choose to communicate during synchronous consumption, and would this vary depending on the media type being consumed?

Thirdly, with respect to how couples communicated, for four of the five couples Text chat was significantly preferred. VMC was used sparingly, with users typically employing this high-fidelity modality at the start or end of CastAway sessions and relying predominantly on textual communication. However, there are likely a number of factors contributing to this reliance on Text communications, some of which may be specific to this particular study. Firstly, as discussed in subsection 4.1, the use of a smartphone for facilitating communication may have biased these results. Secondly, Text is likely better suited to media multitasking (as suggested by our participant's frequent opening/closing of the text entry dialog), allowing couples to attend to the TV content with relatively little audio/visual interruption, and control the attentional demand of communication as compared to media consumption. Thirdly, there may still be social acceptability issues around VMC use, given only half of participants suggested they regularly used VMC with their partner prior to this study. As such, this result will require further experimentation to validate.

RQ 1.4. When do users perceive the need to re-synchronize with their partner (if any)?

Finally, the reliance on Text likely had implications for RQ 1.4, regarding the extent to which users perceived the need to re-synchronize their streams. As demonstrated in the results section, there were a number of occasions across the couples where Chromecast playback de-synchronized significantly (e.g. multiple caching events, internet connectivity problems). However, the number of re-synchronization attempts was low, and often significant delays were tolerated by users, with seemingly little effect on perceived synchronization for the system as a whole. If the couples had relied upon VMC, and perceived the crosstalk of hearing their partners TV playback in the background, we would suggest that there would have been significantly more re-synchronization attempts, in line with [Geerts et al. 2011]. However, because the couples relied on Text communication, delays in content playback were likely made less perceptible, or, at the very least, not important enough to merit an attempt to re-sync.

4.3. Design Implications and Future Research

These results provide a number of considerations for the future design of smartphone and TV-based synchronous at-a-distance experiences, regarding how synchronous sessions are established, communication is facilitated, synchronization maintained, what content is consumed and to whom the burden falls regarding providing such functionality to consumers.

4.3.1. Initiating and Joining Sessions. The majority of CastAway sessions were either prescheduled, or arranged ad-hoc, with only 9 sessions initiated through prompts by the device regarding the partner's availability. This is likely reflective of the lack of usefulness of such a feature given only one potential partner with whom daily contact occurs. However, it may also be indicative of deficiencies regarding how knowledge of a partners availability is noted, and awareness conveyed. For example, due to technical limitations, there was no way to facilitate the joining of on-going Chromecast sessions. Thus, we could not communicate awareness of activity more casually e.g. showing that a partner is watching or listening to content on their TV currently. This usage could also be reflective of a preference toward scheduling and organising these activities as a couples activity or date. Thus, future work will be required in order to investigate how to communicate availability regarding synchronous activity, and how to scale support up to more than two users in one session, allowing for the possibility of serendipitous occurrences of synchronous media consumption.

4.3.2. Supporting Transitions Between Communication Modalities. With respect to communications around synchronous consumption, whilst there was a significant bias toward Text chat, VMC was still utilized by participants, and this bias may have been a result of the particular constraints of this study (see subsection 4.1). However, qualitative evidence suggested that what was most beneficial to users was not any one particular communications channel or modality, but instead the concept of having an open channel, linked to the existence of the CastAway session, upon which text, video and audio chat could be sent and received on demand.

Providing users with the ability to quickly transition between different communication modalities allowed for communication to dynamically vary based on the user's engagement with both the media, and their partner. For example, users were observed to escalate to the higher fidelity modalities (i.e. VMC) when choosing a particular program, then de-escalate to the lower fidelity, lower engagement modality of textual communication for communicating during the program. This was made possible by having the overarching concept of a CastAway session, meaning there was an explicit and accepted link between the participants. This allowed for communications to move away from the concept of calls i.e. explicitly requesting, and accepting, the opening of a two-way communication channel where both sides are continuously broadcasting. Instead, users made themselves implicitly available to receive, and were able to broadcast on any given modality, regardless of their partners choice. Moreover, the always on-top video feed proved popular with users, as it allowed for multi-tasking (predominantly browsing available media to play at the start of a session on the device) whilst still engaging with their partner.

However, such a design comes with some notable problems which were encountered by participants. For example, one couple noted that a partner was using the audio capability to talk when someone else had entered the room, with the user having to mute the phone temporarily. This was problematic because the user could not mute or block a specific communication channel, and the partner communicating has no knowledge that their communication was not being received. Accordingly, more control should be given to the receiver to explicitly block or mute channels when necessary, and these actions should be communicated to the partner, to prevent the frustration of synchronous communication that may not ever be received. Moreover, the utility of this design would need to be evaluated carefully were it to be transposed to a different communication technology, with some of the key benefits (e.g. multi-tasking) likely to be diminished or rendered irrelevant if a different medium than a smartphone were employed.

4.3.3. Synchronization and Control. Although all our participants had high-speed broadband, the variability in terms of internet connection quality was such that significant de-synchronization occurred frequently, with differences of 10 seconds or more instrumented and, surprisingly, largely tolerated, as evidenced by interviews (see subsubsection 3.7.2), questionnaire responses to perceived synchronization and the relative lack of explicit re-synchronization commands (see subsubsection 3.9.8).

By pushing control of re-synchronization to users, we managed to avoid the frequent pauses that would have had to have been enacted for both partners had a tighter synchronization been maintained. However, this observed tolerance goes significantly beyond that noted by [Geerts et al. 2011], where a lab-based evaluation of perceived video synchronization across text and voice chat factors found that for voice chat differences were noticed above 2 seconds difference, and for text chat differences were noticed above 4 seconds.

Whilst these results suggest that this tolerance can be increased, we would caution that this may apply only to specific usage. Firstly, there is the attention being paid to video to be considered. In our study, this attentional engagement varied significantly, as evidenced by interview responses indicating multi-tasking. In contrast, in the study by Geerts *et al.* participants attended solely to the content, in a less relaxed lab environment. Perceived differences in synchronization may have diminished in our study due to this. Secondly, the participants in Geerts *et al.* were made to discuss the content being consumed through a targeted quiz based on the current content. In contrast, in our study participants were free to discuss what they wished during consumption – if their conversations lacked detail regarding what was being displayed, then they may never have become aware of these synchronization differences, regardless of communication modality. Thirdly, our participants predominantly relied on textual communication, which Geerts *et al.* demonstrated was more tolerant to synchronization delays.

On this basis, we would suggest that these findings be carefully considered when implementing automatic re-synchronization, in order to avoid unnecessary disruptive re-synchronization of playback. More relaxed constraints appear reasonable on the basis of these results, however likely only for specific scenarios, which will require further research to establish. In comparing our study to Geerts *et al.* it would appear that constraints likely vary across content genre, the timeliness of the content (e.g. live sports), the engagement of the users (e.g. becoming less necessary when the TV fulfils the function of shared background noise), the communication modality employed (e.g. textual versus audio chat) and the extent to which the content is the focus of the conversation. In such cases, avoiding unnecessary re-synchronization may be more beneficial to users than aggressively maintaining synchronization.

Giving both users complete control over the casting session also proved valuable according to the qualitative findings, allowing users to pause playback when otherwise engaged, implicitly informing their partner that they were unavailable in the process. However, preventing unwanted actions, and having the ability to decouple or relax synchronization (for example if a user wishes to re-watch a part of the show they missed) may merit investigation, although we did not encounter any evidence as to this need here.

4.3.4. Suitability of Non-Video Content To At-A-Distance Consumption. Consuming TV content (video) synchronously at-a-distance represented the most frequently occurring use case, but other non-interactive media forms may merit consideration. Whilst only two couples out of five synchronously consumed Music together, for them it provided a shared backdrop that was compatible with other activities (e.g. tidying) with lower attentional demands than TV content, facilitating connectedness and supporting gift giving behaviour (e.g. selecting tracks for a partner). Advances in network-connected home audio systems are making it feasible to create shared, synchronous sound spaces at-a-distance. Whilst Music in this study was played back on the TV only, for couples that are separated at-a-distance, having a shared background of Music upon which communication is also facilitated appears capable of playing a similar, yet distinct, role compared to TV content, in terms of facilitating connectedness for the subset of couples for whom consuming Music together is appropriate (e.g. those with aligned tastes in music).

4.3.5. Content Providers and Smart TVs: Who Should Provide At-A-Distance Functionality?. Users repeatedly noted that the limitations in terms of content platforms (BBC iPlayer and Google Play Music) impeded their capability to select content to view together. Thus, relying on content providers to independently implement and support synchronous at-a-distance functionality on their particular platforms is problematic, as this will lead to both an increased cost to the content providers, and frustration for the users when any given platform does not support shared experiences at-a-distance.

Accordingly, we suggest that the smart TV platforms should bear the burden of supporting synchronous at-a-distance playback of media content. By this we mean that the smart TV platform should support multi-user, multi-device control and the synchronization of content playback state across multiple playback devices, with said support built into the playback and media APIs used by content platform applications. In doing so, a consistent experience can be provided, agnostic of content provider applications, and support provided to users regarding the discovery and joining of on-going consumption activity across a range of content providers, much as how multi-player gaming is currently facilitated (e.g. unified friends lists, joining on-going games etc.). CastAway serves as a demonstration of the technological feasibility of such an approach.

There are, however, a number of issues that would need to be resolved for a smart TV platform to provide this functionality, from engineering to user experience. For example, further investigation will be required to determine the constraints regarding synchronization, and the distance over which consumption is feasible (e.g. will synchronization issues compounded by the latency of communications between the UK and Australia prevent separated friends or family from effectively viewing together?). Integrating communications functionality puts an additional burden on the smart TV platform, and may have implications regarding privacy if there is a concerted push toward VMC facilitated by cameras embedded in the living-room).

Digital rights management may prove problematic e.g. when one user has access to the UK-based library and another has access to an Australian-based library. And supporting scalability across more than two people or more than two households (e.g. in terms of how VMC is facilitated [Williams et al. 2011]) would have significant implications for synchronization, communication and interaction. Nonetheless, as an exemplar the approach taken by CastAway offers a low barrier of entry for consumers and smart TV platforms, given the ubiquity of smartphones, the popularity and low cost of Google Chromecast and the availability of Chromecast support in Android TV, a smart TV Operating System adopted by a number of TV manufacturers (Sony, Sharp and Phillips).

5. STUDY 2: THE ROLE OF VR HMDS IN AT-A-DISTANCE MEDIA CONSUMPTION

CastAway investigated TV at-a-distance from the perspective of what is currently possible with existing consumer technology. By this we mean that social presence was conveyed through a smartphone with a standard messaging interface supporting text, audio and video. And immersion was provided through the home TV, typically the biggest display in the room. As such, the applicability of CastAway is, to a large extent, predicated upon the assumptions that text/audio/video messaging will remain the preferred means of conveying social presence and communication, and that the TV will remain the most immersive and enjoyable display in the room on which to consume entertainment content. However, the imminent availability of consumer VR HMDs (e.g. Oculus Rift, Sony Playstation VR) has the capability to undermine these assumptions.

Firstly, VR HMDs empirically provide a level of immersion (i.e. place and plausibility illusion) which significantly goes beyond what a typical TV is capable of, with a range of new film (e.g. immersive 360° video [Disney 2015]), TV and gaming media being created to take advantage of this fact. The consumption of traditional 16:9 TV content could also be affected, with additional immersion potentially being induced through viewing the content in your own personal cinema, or augmenting the virtual environment to match said content.

Secondly, combined with appropriate sensing technology and a network connection, VR HMDs also have the ability to enable telepresence, where those at-a-distance could appear to share a virtual space with a VR HMD user. By assuming the presence of room-wide sensing (e.g. RGBD cameras such as the Kinect V2 mounted in a space such that they have a view of the room) and head-mounted sensing (e.g. wide angle cameras such as the Leap motion attached to the front of a VR HMD) mixed reality experiences using VR HMDs become feasible. For example, prior telepresence research has seen users at different remote locations captured via cameras, with this captured imagery transmitted over a network and rendered in-place in mixed reality [Billinghurst et al. 2002]. Beyond rendering users in PiP directly on the TV, when combined with VR / AR HMDs this also opens up new possibilities for users at-a-distance to share virtual spaces, without relying on virtual avatars [Greenhalgh and Benford 1995] or telerobotics[21]. Instead, this social presence could potentially be communicated as in reality, complete with position, gestures, facial expressions, and clothing. In the home, these VR HMDs will, in some cases, find themselves connected to powerful, networked, RGBD-equipped games consoles (e.g. PS4, XBOX One) which already feature prominently in the living room.

In effect, the assumptions underpinning the creation and evaluation of CastAway regarding immersion and communication may, in the near future, be undermined to some unknown extent. In accepting this, there are implications regarding synchronous at-a-distance media consumption that remain as-yet unexplored. Thus, our second aim was to prototype and evaluate VR HMD-based at-a-distance experiences, such that we begin to explore the longer-term role of both the TV, and VR HMDs, in at-a-distance media consumption. Consequently, we formulated three research questions pertaining to the application of VR HMDs to synchronous at-a-distance media experiences:

RQ 2.1. To what extent can VR HMDs approximate the experience had when physically co-viewing together?

RQ 2.2. To what extent can VR HMDs exceed the experience had when co-viewing together using the TV at-a-distance?

[21] http://www.engadget.com/2015/10/26/oculus-rift-teddy-bear-adawarp/

RQ 2.3. Will the media immersion provided by VR experiences help or hinder socialization at-a-distance?

5.1. Design and Demographics

Five Conditions were defined to answer these research questions (see Figure 18), comparing TV together and at-a-distance with VR HMD use at-a-distance, with support for embodied telepresence, and varying levels of media immersion. For the baseline Condition, participants were physically seated side-by-side with shared viewing of a TV, whilst for the at-a-distance Conditions, participants were seated at opposite ends of the lab, wearing audio headsets with microphones. In the VR Conditions, participants could see their partner as if seated next to them, just outside the peripheral view when looking straight ahead in each scene, with one participant captured from the left, one from the right such that participants appeared in the correct orientation to their partners. In all Conditions, awareness of partners was kept approximately the same i.e. full body actions and gestures were capable of being observed. Participants were seated throughout. The Conditions were:

1: TV Together. This was the baseline for viewing together, with participants seated next to each other viewing a $24''$ TV display.
2: TV at-a-distance. This was the baseline for viewing apart, based on the preference for PiP in the field study from [Macaranas et al. 2013], with participants being able to see each other in the bottom left/right hand corner of a $24''$ TV.
3: VR TV at-a-distance. Here participants found themselves in a photosphere of the same room from Conditions 1 and 2, being able to see their partner sitting to their left/right, viewing media content on a virtual screen of similar size to that in Conditions 1 and 2.
4: VR Cinema at-a-distance. Here participants were in a 3D virtual Cinema scene, with media content playing on a Cinema-sized virtual display.
5: VR 360° video at-a-distance. Here participants found themselves in 360° video sphere scene.

For all Conditions bar Condition 5, the media content comprised of 8 minute clips from a nature documentary series [Attenborough 2006]. For Condition 5, the content comprised of 360° nature documentary clips from [Discovery 2015]. These Conditions were chosen to quantify how much closer a person might feel to their partner when watching media content sitting next to them in a virtual space, compared to prior VMC approaches (comparing Conditions 2 and 3, answering RQ 2.2) and physically sitting together (comparing Conditions 1 and 3, answering RQ 2.1). Secondly, this design allowed us to examine the effect that increasing the immersion in the media content had on participants' capability to attend to, and communicate with, their partners (answering RQ 2.3). It was intended that Condition 3 would offer the lowest immersion in the media content, owing to the dimensions of the virtual TV, and the context of the virtual setting (the lab space the participants were physically in, portrayed via photosphere). The setting of Condition 4 is that which is typically considered most immersive for widescreen content, a Cinema. Condition 5 featured recorded 360° video, allowing the viewer to become encapsulated in the recorded world in the most immersive media content. In this way, Immersion was controlled by varying the environment and the size of the display (audio was consistent throughout). Conditions 2-5 were counterbalanced, with Condition 1 (baseline) always recorded first to get an accurate baseline prior to VR Conditions. Each Condition lasted 8 minutes. Video clips were not randomized due to the limited amount of comparable 360°footage available for Condition 5. As such, clips were vetted for similarity in terms of content and narration. Participants were recruited from University mailing lists in pairs that knew each other, with 12

1: TV Together, with participants seated side-by-side in reality

2: TV At-A-Distance, with participants able to communicate via PiP video and headphones/microphone

3: VR TV At-A-Distance, with participants wearing VR HMDs and headphones/microphone, able to see and hear each other in VR, set in a photosphere of the real-world lab setting

4: VR Cinema At-A-Distance, as with (3), except set in a VR cinema

5: VR 360° At-A-Distance, as with (3), except set in a 360° VR video experience

Fig. 18: Conditions for Study 2. Condition 2 shows the view on one participant's TV screen in reality, Conditions 3 to 5 show views from the perspective of a VR partner. In all Conditions the viewing is from the perspective of the rightmost partner.

pairs (24 participants, 18 males, 6 females, 3 pairs in relationships, 9 friendship pairs) recruited, with an average age of 21.6 years (std.dev=4.2).

Through these conditions, shared immersive experiences are compared to physically co-viewing together, and co-viewing together at-a-distance using the TV for both viewing and communication, to answer RQs 2.1 and 2.2. Three different levels of media immersion are investigated, and socialization examined both quantitatively (through the amount of speech, and the amount of time spent looking at a partner) and qualitatively (through questionnaires, see "Measures") to answer RQ 2.3.

5.2. Measures

After each condition, a 36-item questionnaire was delivered to participants. This comprised of a 20-item TV / media immersion scale (from [du Toit 2012], used in [Brumby et al. 2014], derived from questionnaires for immersion in gaming [Jennett et al. 2008] and narrative engagement [Busselle and Bilandzic 2009]), Social Presence and Closeness [Macaranas et al. 2013], emotional connection [Forghani et al. 2014], synchronization [Geerts et al. 2011] and engagement/togetherness [Inkpen et al. 2013]. The TV / media immersion scale was employed because it offered a universal measure of gauging immersion in a media experience, across both TV and VR conditions, that was not biased toward constructs of VR HMD-based presence and immersion specifically. Finally participants ranked the at-a-distance Conditions in order of preference, with short interviews conducted regarding preferences. Across all conditions, duration of participants speech was recorded, whilst for the VR Conditions the azimuth and polar angles of viewing were also recorded at 20Hz as a means of measuring workload and engagement. See Appendix B for the questionnaire. Friedman tests with *post hoc* Bonferroni corrected Wilcoxon tests were conducted where applicable.

5.3. Implementation

For implementation of these conditions, we used consumer VR HMDs (2 * Oculus Rift DK2, SDK v0.7.0). At a resolution of $960 * 1080$ per eye, 1080p media content effectively loses half the horizontal resolution when in full view (e.g. Condition 4), appearing as SD TV content in Condition 3. Whilst there is a loss in clarity, pictures were eminently viewable on the best consumer grade VR HMD available at the time. We built a system that allowed for pairs of users to engage in shared VR experiences using the Unity 3D engine (v5), with the capability to see and hear each other in these experiences as if seated next to one another, by using Microsoft Kinect V2s, and audio headsets with microphones, based upon the code made available by [McGill et al. 2015a]. This system allowed us to present 360° photos, videos (using the Renderheads AVPro library [Renderheads 2015] for high-performance video decoding), and fully virtual 3D content in synchronization across our users, locked at 75FPS.

5.4. Results

5.4.1. Questionnaire (RQ 2.1 / RQ 2.2). For the questionnaires there was no significant effect on Social Presence, with the VR social presence scores broadly comparable with the Control. There was however a significant effect on immersion, with *post hoc* tests validating that Conditions 4 and 5 were the most immersive in terms of media content, compared to Conditions 1–3 (see Figure 19). Whilst it is surprising that the Control condition (seated side-by-side) is not significantly greater in terms of social presence than the computer-mediated communication Conditions, this result is likely reflective of social awareness provided by each Condition. Whilst participants were seated side-by-side, they were under no obligation to attend to each other, other than if they wished to observe each other's reactions and expressions, requiring a transition in gaze (i.e. looking to the left/right). In addition, with the TV acting as a singular, central focus of

attention, there was little opportunity whilst consuming TV-based media for serendipitous viewing actions to occur between participants.

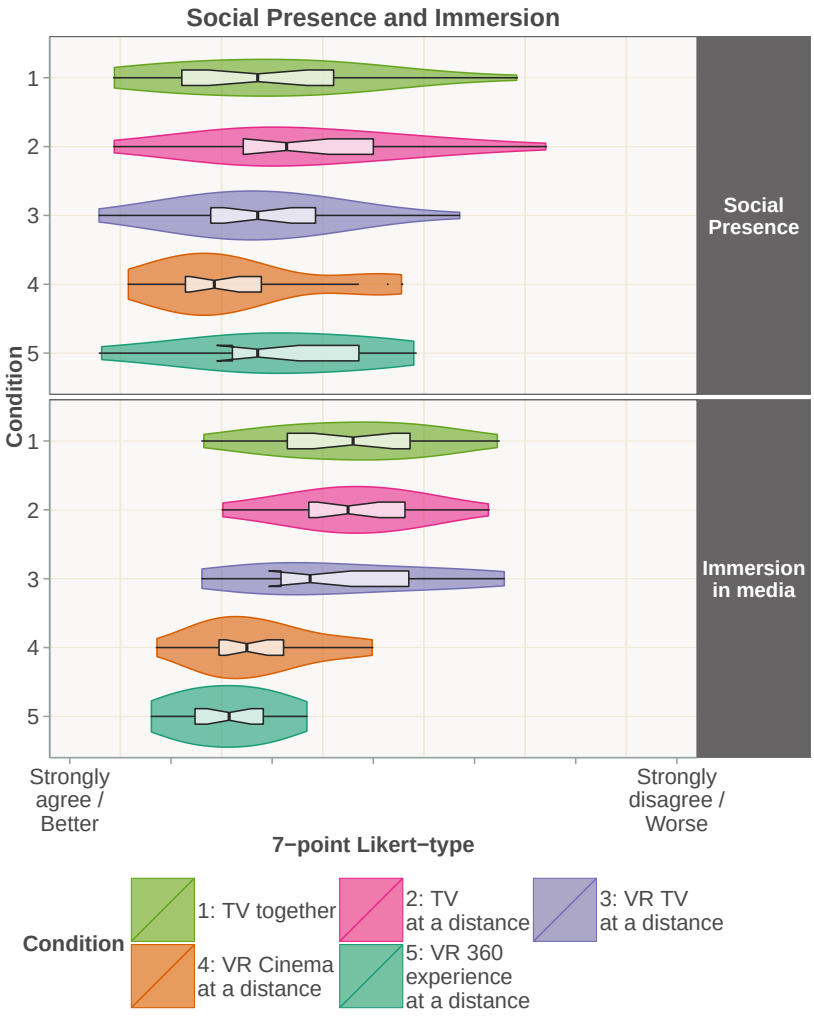


Fig. 19: Scores from Social Presence [Macaranas et al. 2013] and Immersion [du Toit 2012] questionnaires. **Social Presence**: $\chi^2(4) = 6, p = 0.2$. **Immersion**: $\chi^2(4) = 58, p < 0.01$, *post hoc*: 1-4, 1-5, 2-4, 2-5, 3-4, 3-5

In contrast, in Condition 2 (at-a-distance TV) social observation is made casually accessible by the fact that it can be accomplished by a small eye-based gaze transition, as the partner is viewed on the TV alongside the content. Condition 3 (VR TV at-a-distance) is largely the same as Condition 1, which is to be expected given they are essentially the same environment. Conditions 4 and 5 (VR Cinema and 360 experience) both feature VR environments which encourage visual 360° exploration (to

different extents), meaning that there is a greater likelihood of inadvertent observations and serendipitous interactions with the partner. In this way, the questionnaire results appear to accurately reflect the differences in terms of how social information is conveyed and made accessible across the Conditions.

The extent to which participants felt engaged and enjoyed the experience both featured significant effects, with the more immersive VR conditions being more engaging and enjoyable. Moreover, Condition 5 (360° experience) is perceived as being more enjoyable than Condition 4 (VR Cinema). There was a significant effect on the emotional scale but *post hoc* tests found no significant differences (see Figure 20). Condition 2 (at-a-distance TV) continues to come out worst on all scales.

For synchronization there was a significant effect between Conditions 2-5. This suggests that the immersion provided by Condition 5 impacted perceived synchronization, and that participants were, in part, responding to the synchronization question not from the perspective of technological synchronization, but experiential synchronization. For togetherness there were no significant *post hoc* differences, whilst for experience there were significant *post hoc* differences between 2-4 and 2-5 (see Figure 21).

The rankings largely mirrored the rest of the questionnaire results (see Figure 22). Firstly, the VR conditions were preferred to TV at-a-distance, indicating that the virtual room environment and embodied communications of Condition 3 was preferable to PiP TV-based at-a-distance communications. The more immersive VR conditions were in turn preferred to the VR room environment of Condition 3, being more engaging and enjoyable. Embodied VR communication was preferred to video-mediated communication (Condition 2 versus 3) whilst more immersion was also preferred (Condition 4/5 versus 3). However, there was no significant difference between Conditions 4 and 5 - whilst condition 5 was more enjoyable, both experiences had significant merit.

Responses to a post-study question regarding the likelihood of use of VR (see Figure 23) indicate that having the capability to socialize in the manner experienced, through mixed reality communication, was a motivator for adoption of VR use for media consumption (mean=2.21, std.dev=1.35).

5.4.2. Viewing Activity (VR Conditions only) (RQ 2.3). Participants' viewing was instrumented, with gaze orientation (Oculus Camera orientation in the 3D engine, linked to the positional / rotational orientation of the HMD) recorded at 20Hz during the VR conditions, to establish the effort expended in viewing, and the extent to which participants looked at their partner, the virtual media screen (if applicable), and their VR environment.

Participants looked in the direction of their partner (see Figure 24) the most in Condition 5 (VR 360°) ($\chi^2(3) = 17, p < 0.01$, *post hoc:* 3-5, 4-5), on average for 86 secs (std.dev=32 secs) compared to 47 secs (std.dev=39 secs) for Condition 4 (VR Cinema), and 51 secs (std.dev=41 secs) for Condition 3 (VR TV). Condition 5 featured the most varied viewing of the 360° space, owing to the immersion and novelty of the 360°video. This, in turn, also caused the most head movement (see Figure 25), as measured by cumulative great circle distance[22], meaning that Condition 5 also elicited the most physical effort from participants.

Regarding the amount of effort expended, as measured by the great circle distance of gaze changes, in the VR conditions over time (see Figure 26), it can be seen that Conditions 3 and 4 feature approximately the same amount of effort, with relatively little degradation in that effort over the 8 minutes of viewing in each Condition. In contrast,

[22]The great circle distance is the shortest distance between two points across the surface of a sphere. By iterating through the recorded gaze logs and summing the great circle distance for the unit sphere between the current point of view and the previous point, a measure of the total head movement is established.

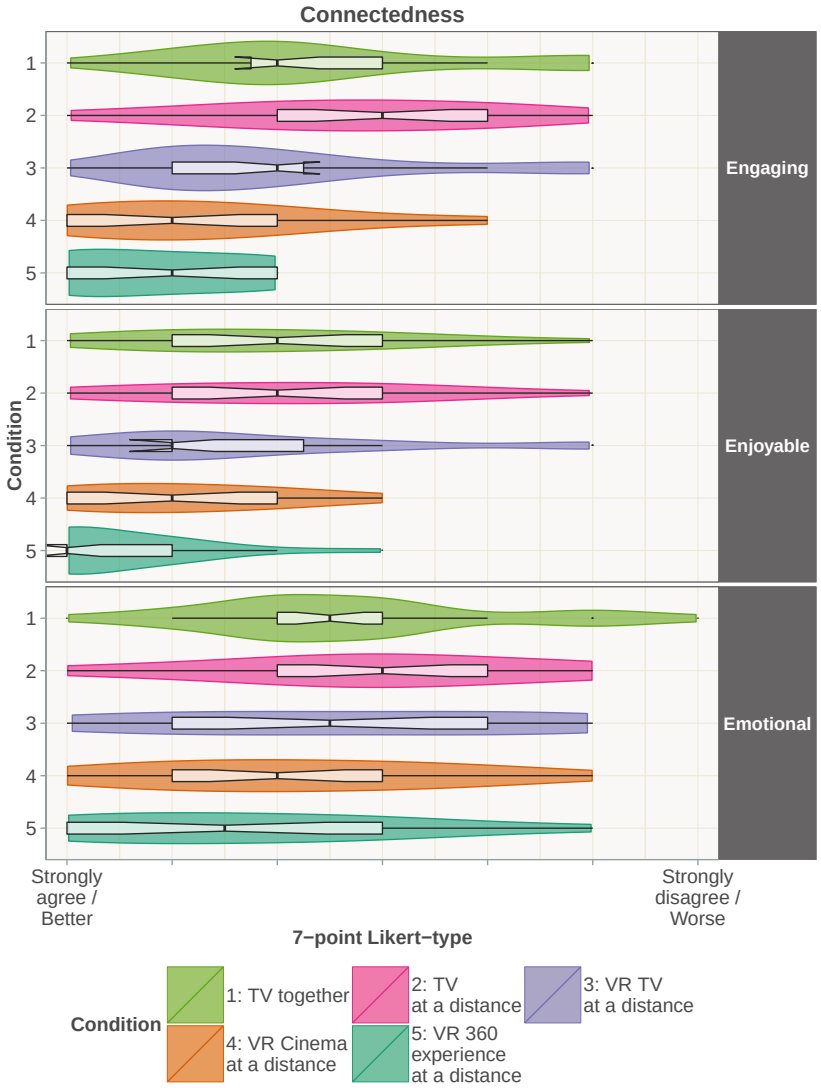


Fig. 20: Responses for Connectedness questions [Forghani et al. 2014]: "How engaging / enjoyable / emotional was it when you viewed media content with your partner?". **Engaging:** $\chi^2(4) = 46, p < 0.01$, *post hoc:* 1-4, 1-5, 2-4, 2-5, 3-4, 3-5. **Enjoyable:** $\chi^2(4) = 38, p < 0.01$, *post hoc:* 1-4, 1-5, 2-4, 2-5, 3-4, 3-5, 4-5. **Emotional:** $\chi^2(4) = 16, p < 0.01$), *post hoc* tests found no significant differences.

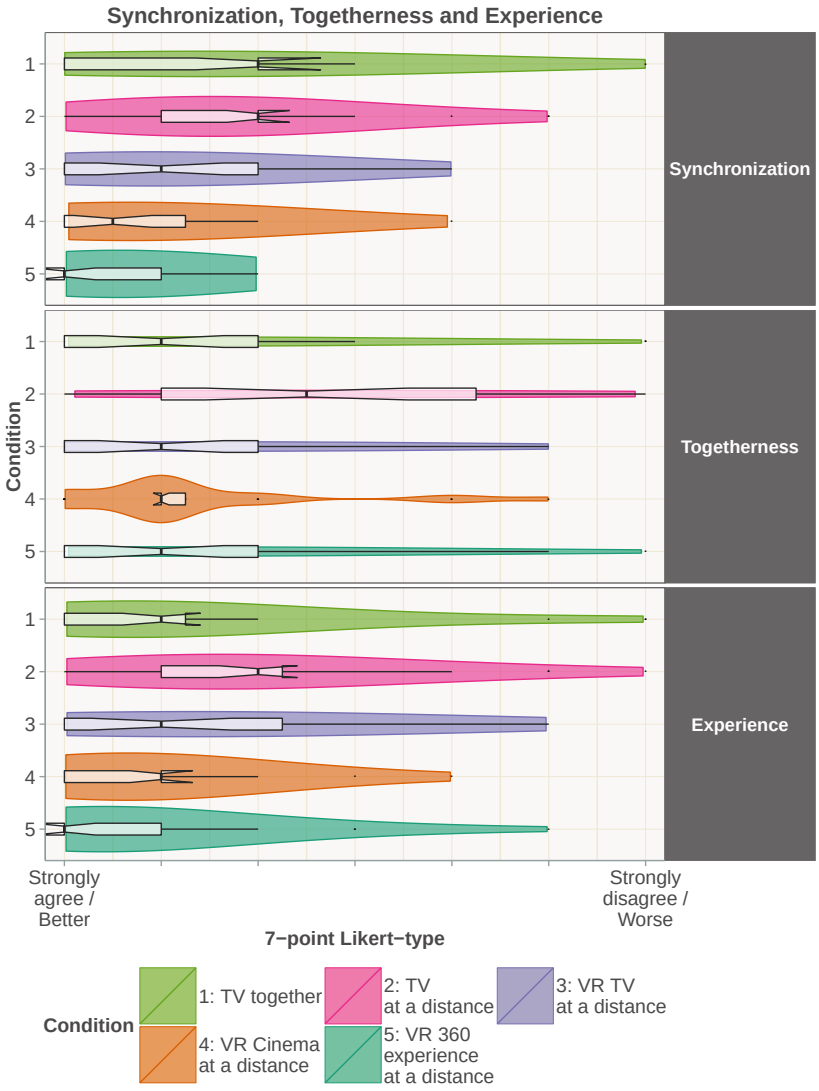


Fig. 21: Responses for Synchronization [Geerts et al. 2011] ("It felt like our experiences were synchronized", Togetherness ("It felt like my partner and I were interacting naturally, as if we were together" and Experience ("It felt like I was experiencing the activity with my partner" [Inkpen et al. 2013]. **Synchronization:** $\chi^2(4) = 15, p < 0.01$), *post hoc* 2-5. **Togetherness:** $\chi^2(4) = 10, p < 0.05$), *post hoc* there were no significant differences). **Experience:** $\chi^2(4) = 16, p < 0.01$, *post hoc* 2-4 and 2-5.

Condition 5 features a marked increase in effort expended, with effort decreasing over time, and increasing at the advent of a novel, new VR clip. Clip 3 featured a marked increase in effort in comparison to Clips 1 and 2, likely due to the fact that this clip featured significant continuous on-going activity occurring within the full 360° scope of the video (in the form of a number of sharks swimming around the viewer concurrently), in contrast to Clip 1 which featured 360° activity to a lesser degree (occasional sightings of fish), and Clip 2 which primarily featured a single focus of activity (viewing a person engaged in nature activities). These findings suggest that, for short durations

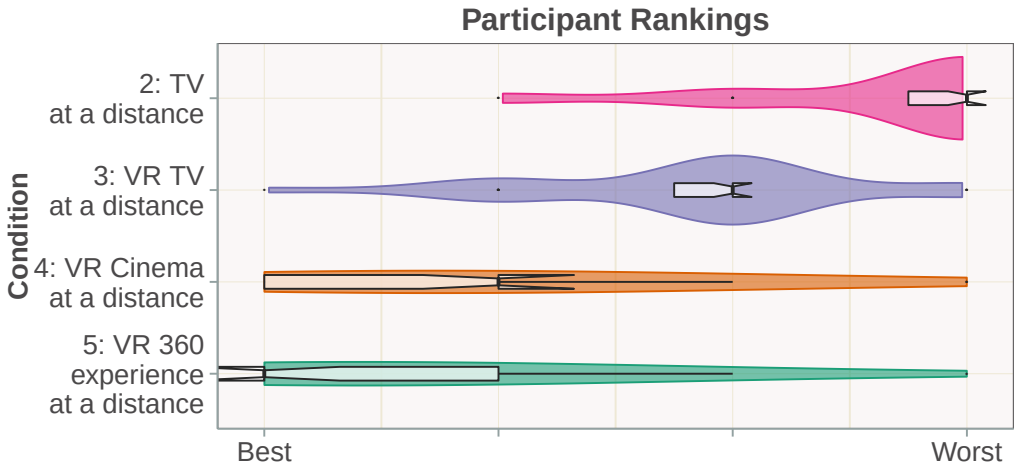


Fig. 22: Rankings (lower is better) for Conditions 2-5. $\chi^2(3) = 38, p < 0.01$, *post hoc:* 2-3, 2-4, 2-5, 3-5).

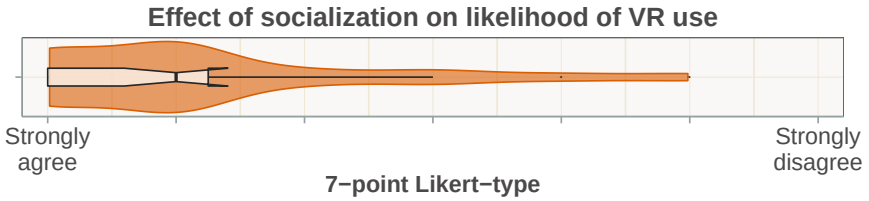


Fig. 23: Responses to "If more people could be brought into the VR environment at a distance (e.g. watching content together with friends or family), this would make me more likely to consume media in VR."

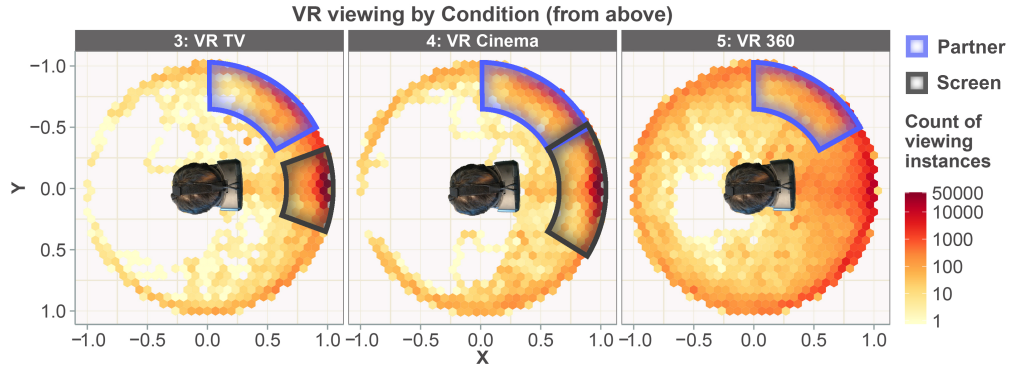


Fig. 24: Hex bin plot of total viewing instances across participants (as seen from above, with straight ahead for the participant at the rightmost point of the circle, as shown by HMD wearer at (0,0)). Viewing was sampled at 20Hz, meaning each viewing instance approximately accounts for 50 milliseconds of viewing. Note the log scale for the count of viewing instances.

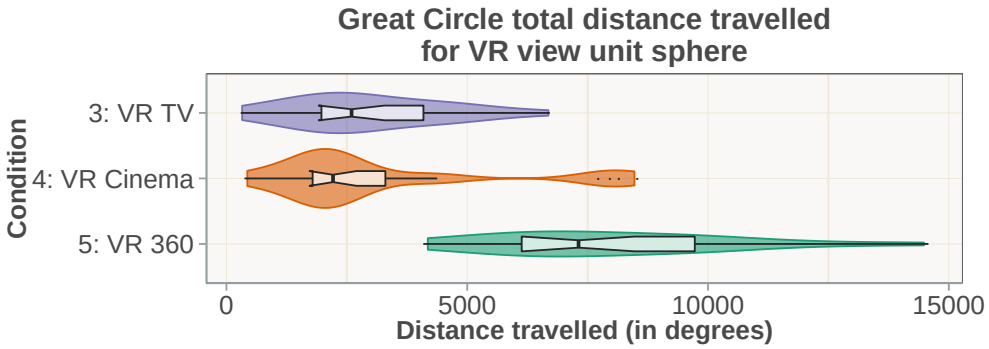


Fig. 25: Cumulative Great Circle distance between sampled viewing points for VR view unit sphere (rho=1), across participants. $\chi^2(2) = 28, p < 0.01$, *post hoc:* 3-5, 4-5.

at least, as the amount of activity available to attend to in a 360° experience increases, so too will the effort expended in attempting to view said activity. Whilst fatigue may play a role here, as suggested by the decline in viewing activity in each clip, this decline could also be due to the increasing familiarity and thus decreasing novelty of the clips over time.

5.4.3. Speech (RQ 2.3). While there was a significant main effect for amount of speech (see Figure 27), *post hoc* tests revealed no significant differences between the Conditions, with Condition 5 featuring the highest mean speech of all the Conditions.

5.4.4. Interviews. As with Study 1, interview transcripts were first coded using an Initial Coding approach, then grouped using a thematic approach and analysed based on the frequency and importance of the codes. Two themes were identified: *Effect on Socialization* and *Media Preferences For At-A-Distance Consumption*. Excerpts are broken down by pair, with each participant discussing the theme quoted.

Effect on Socialization. This theme encapsulated codes regarding "interaction with partner", "awareness of partner" and "acceptability". One participant pointed out that their engagement with their partner might not be wholly reflected in viewing data:

> *P2*: I probably looked to him less during the 3D one but we were possibly interacting more because we were saying "aww, look behind you" and stuff, so there's more interaction there as compared to sitting watching TV.

The unfamiliarity of socializing via VR proved difficult for two participants. Whilst participants could approximately perceive their partner as sitting next to them, the underlying knowledge that they were not actually there affected their reactions and treatment of their partner, despite of knowing their partner could perceive them in the same fashion:

> *P22*: It's a bit strange to think you're sitting next to someone, and then like ah, you're not actually sitting next to each other. But I mean maybe it's like Skyping. If you get used to it, then it's like you know that you're not next to them.
>
> P8: Sometimes if I'm talking to you I look across, but I don't ever look across to check you're there. Because physically you're not there, I don't feel I need to probably do that.

A lack of peripheral awareness of the partner's activity, a technical limitation due to the 110° field of view of the VR HMD, impeded socialization for two participants by

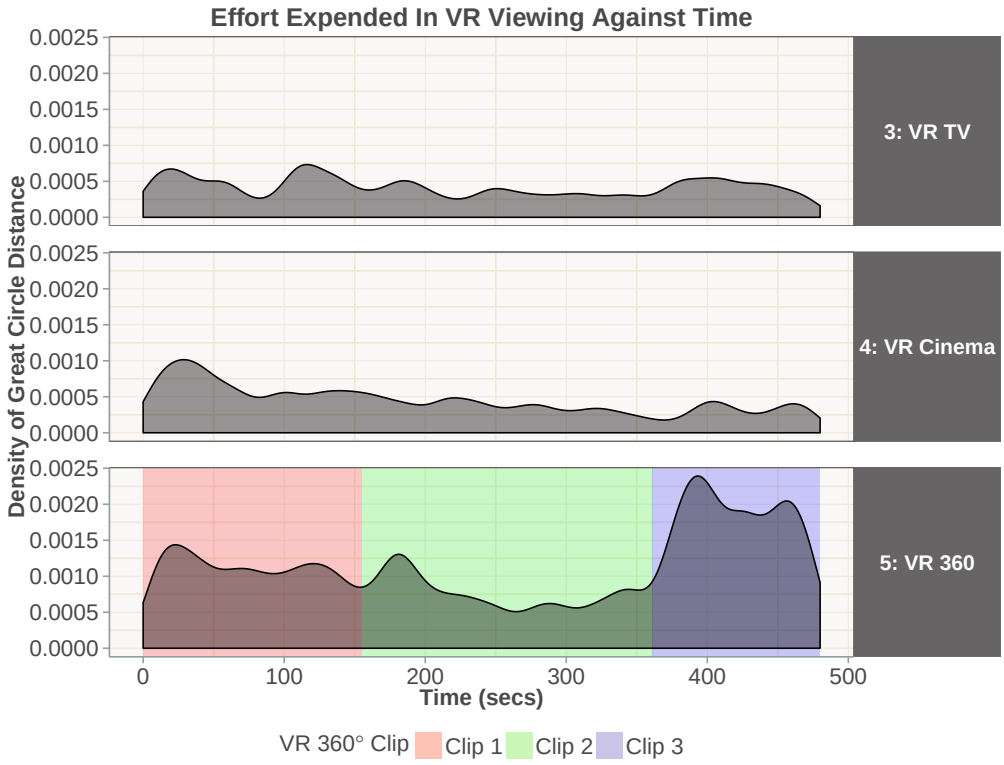


Fig. 26: PDF of effort expended in VR viewing (great circle distance) over time by Condition. For Condition 5, three clips were used, transitions between these clips are labeled.

diminishing the sense that their partner was actually presence. However, for one user this lack of peripheral awareness meant that a more explicit decision to attend to their partner had to be made, increasing their connectedness.

> *P1*: One thing I didn't like, I wasn't able to see (my partner) at all, I had to move my head... it felt like he wasn't there for me.
> *P2*: You have to turn deliberately, you don't see any of them, they're still either just there and you can see if they gesticulate, or they're not there at all until you move your head to the side.
> *P1*: Because when you are on the sofa with your peripheral vision you can see what's happening.
>
> *P24*: I liked how in all of them you had to turn round to look at them to see, so you could just choose whether you wanted to (see your partner).

For at least one participant the Cinema context inhibited socialization due to the societal norms attached with viewing in such spaces (e.g. where talking is frowned upon):

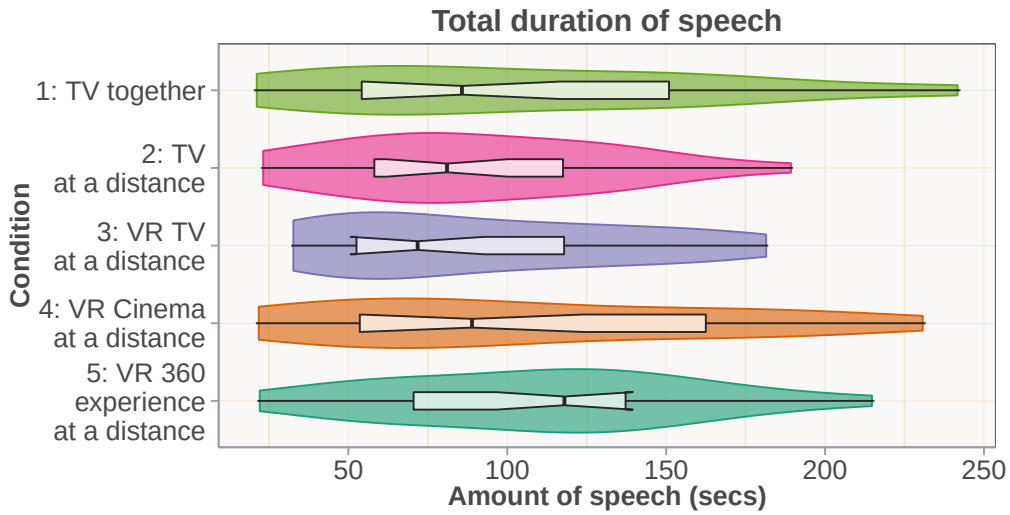


Fig. 27: Total duration of speech (seconds) across participants. $\chi^2(4) = 11, p < 0.05$, *post hoc* tests revealed no significant differences.

> *P11*: For socializing, I'd say [Condition 3 was the best], I think the cinema is the sort of environment you go there to just to watch things, you don't really socialize, whereas the TV was the more social environment.

Regarding RQs 2.1 (approximating experience when physically together) and 2.2 (exceeding TV at-a-distance), the VR Cinema proved popular for consuming TV media in user rankings. This was reflected in comments from four participants, which suggested that recreating a familiar space associated with viewing together made the experience appear more real:

> *P1*: I liked the movie theatre because it was like a physical link together... I felt like I was in the same room as (my partner).
>
> *P9*: Honestly, I'd watch stuff like that. *P10*: It's far better than watching it in (Condition 3)
>
> *P12*: It gets the experience of being in a Cinema, it felt the same. Like, dark and... If I had popcorn then it'd be good for it.

Media Preferences For At-A-Distance Consumption. With 360° video, the majority of participants noted their enjoyment of said content, noting the immersion and sense of presence provided compared to TV content:

> *P11*: 360° video I really liked, I felt like you were actually there. I've heard the term presence thrown around in VR and I'm starting to kind of understand what that is now, because you kind of feel like you are actually there rather than just watching a TV screen.

However, in terms of utility for at-a-distance use it proved divisive. Five participants suggested that the immersion provided by the 360° experiences impeded their ability to have a shared experience with their partner, for a variety of reasons. The attentional demands of 360° media, the lack of a shared focus of attention or cues to guide a shared

focus, the exertion of consumption and the pressure of trying to ensure that nothing is missed all contributed to anxieties regarding using this form of content:

> *P6*: The 360 one is too immersive, and you don't want to pay attention to your partner, it's too distracting.
>
> *P8*: Sometimes you just don't know where to focus, it's like I don't know if I should be focussing on (my partner) or a shark, or the back or the side. If I don't look somewhere it's like I'm going to miss something... it's just another thing you have to look at. Or one other thing to not look at. I don't know if I liked it or not.
> *P7*: It's different though, it's more active, it was something we'd do for short periods of time where you're looking around you and engaged in what's going on... I think it would be exhausting watching a VR movie where there's also something that could grab your attention, but you don't know where it is, like (my partner) said.
> *P8*: And if you don't look, you've kinda missed it. Whereas if you look at a TV, everything is always in front of you, you can't really miss content.
>
> *P11*: There was too much going on in 360, you couldn't watch a movie in it because you'd be looking about constantly, (whereas) the VR cinema was just there.
>
> *P23*: The 360 one is kind of hard because you wouldn't really watch a lot of TV like that... you'd always feel like you are missing something. It would have to be something like the ones that you showed us... it's just something's happening around you, like an experience.

6. DISCUSSION OF STUDY 2

Based on these findings, we firstly examine limitations regarding the external validity of our findings, before discussing the implications these findings have for our research questions. Finally, we discuss the subsequent design implications and future work that arise from this study.

6.1. Limitations

Interpretation of the findings of this study should take into account limitations regarding the design of the communications functionality in the TV at-a-distance condition, the sample size, population and duration of study which affect external validity and the novelty effect of VR use.

6.1.1. TV At-A-Distance Communications Design. The TV at-a-distance condition (Condition 2) facilitated communications through use of a headset with microphone (for audio) and a picture-in-picture full-body view of the participant (for video). Firstly, it should be noted that this reflects only one of a multitude of ways by which at-a-distance communication can, and has, been facilitated. For example, in terms of the technology used, alternatives include using mobile devices (as chosen in CastAway due to their ecological validity and suitability for deployment) or using separate fixed screens (e.g. off-TV proxy placement, laptop/tablet), as well as considering different camera positions and coverage (e.g. using the in-built camera on a mobile device, versus a camera in the environment). Furthermore, textual chat was not facilitated, with VMC being the only communication modality available.

This design was chosen for a number of reasons. Firstly, it was impractical to consider every permutation of at-a-distance communications functionality (e.g. text/audio/video using mobile devices, fixed screen, and overlayed/picture-in-picture)

in one study. Accordingly, this condition was chosen on the basis of being representative of the findings of previous TV at-a-distance studies (namely [Macaranas et al. 2013]) regarding user preference for picture-in-picture. The particular representation of the user (full body view) was chosen because it kept the potential awareness of the at-a-distance viewer consistent throughout the conditions. This was to guard against fluctuations in social presence if, say, full body gestures were visible in all conditions bar one. Consequently, whilst this study allows for some determination as to how this specific variant of TV at-a-distance is perceived relative to VR at-a-distance, applying the same determination to other variants of TV at-a-distance will require further research.

6.1.2. External Validity. The population sample for this study was made up of friendship pairs (9 of the 12 pairs) and couples (3 of the 12 pairs), with the scope of recruitment widened compared to Study 1, effectively examining pairs with a comparatively low need for at-a-distance experiences (compared to couples at-a-distance). This choice was made firstly because of the difficulties in recruiting sufficient numbers of at-a-distance couples, given this was a collocated laboratory study and thus would require them to be physically together. And secondly because, unlike in Study 1 where usage was dictated by the participants, in this study usage was controlled i.e. participants would be exposed equally to each of the viewing conditions, allowing direct comparison. Whilst this recruitment would have an effect on the external validity of the results (i.e. the extent to which elicited preferences would apply to usage by geographically separated couples or friends at-a-distance), validity in this respect was already compromised due to this being a collocated laboratory study (and not an in-the-wild deployment) by necessity of the technology being utilized. Accordingly, greater consideration was given to having a sufficient number of participants, recruited from demographics that were still relevant and could feasibly be expected to utilize and find value in such systems (e.g. friendship pairs). As such, the external validity of the findings will remain limited until such time as these findings can be validated in an in-the-wild deployment targeting specific demographics (e.g. couples in long-distance relationships, family, friends). However, this study nonetheless featured a high degree of internal validity.

6.1.3. Novelty Effect of VR. Given the duration of each condition, and the specialized nature of the VR HMD experiences, there is a possibility that there was a novelty effect regarding preferences toward VR (although Conditions were counter-balanced). To what extent these findings would hold would need to be examined through either a longitudinal, in-home deployment (e.g. examining Conditions 4/5), or repeated evaluations after further exposure to VR HMDs (i.e. subsequent to consumer adoption). As such, these findings should be considered as tentative initial evidence regarding consumer preferences toward VR at-a-distance, in what is an emerging domain.

6.2. Implications for Research Questions

RQ 2.1. To what extent can VR HMDs approximate the experience had when physically co-viewing together?

With respect to RQ 2.1, there were no statistically significant differences between Condition 1 (baseline) and Condition 3 (VR at-a-distance), suggesting but not confirming that the VR at-a-distance Condition in some ways approximated viewing together physically. However, the similarity between the results of Condition 1 and Condition 2 (TV at-a-distance) with respect to social presence suggests that the measure of social presence used failed to adequately describe the differences between in-person presence and variations in at-a-distance presence, instead measuring functional presence i.e. the user's capability to attend to, and converse with, their partner. Thus, further

measures of social presence should be considered for future studies (e.g. using the TPI [Lombard et al. 2011]).

RQ 2.2. To what extent can VR HMDs exceed the experience had when co-viewing together using the TV at-a-distance?

For communications functionality alone, VR at-a-distance appeared to be preferred to PiP TV at-a-distance by participants. This is evidenced by the differences between Condition 2 (TV at-a-distance) and Condition 3 (VR TV at-a-distance), specifically the participant rankings where a significant difference was found between Conditions 2 and 3. In both conditions the size of the display, and the environment, were controlled for (with Condition 3 effectively being a VR replica of the environment of Condition 2), with the difference between the Conditions being the PiP video of Condition 2, compared to the embodied telepresence of Condition 3.

RQ 2.3. Will the media immersion provided by VR experiences help or hinder socialization at-a-distance?

Regarding RQ 2.3 and the effect of immersion on socialization (considering Conditions 3–5), whilst there was no effect in terms of the amount of speech, there was an effect in terms of the total viewing of partners, with viewing significantly increasing in Condition 5. However, this discrepancy could be accounted for by the fact that participants may not have been intending to look at their partner, but instead their virtual environment. Because of this, the awareness gained through this viewing may be questionable, given that participants may have predominantly been attending to the 360° environment. This is hinted at in the Social Presence scores, with mean Social Presence being lower in Condition 5 compared to Condition 4, but is not proven as the difference was not statistically significant.

With respect to using immersive virtual environments (the cinema of Condition 4) for consuming traditional media content, significant differences were found regarding media immersion, engagement and enjoyment (all improved compared to Conditions 1–3) and experiencing the activity with their partner (improved compared to Condition 2). This suggests that augmenting existing content to increase immersion can have a marked difference on the experience of consuming said content at-a-distance.

Finally, with respect to using more immersive VR media content (the 360° content of Condition 5), this was again preferable with respect to media immersion and engagement (improved compared to Conditions 1–3), experiencing the activity with their partner (improved compared to Condition 2), as well as enjoyment (improved compared to Conditions 1–4) and rankings (preferred compared to Conditions 2 and 3). This indicates that there is some advantage for 360° content, compared to traditional media content that has been augmented to be more immersive (in the case of Condition 4) and traditional media content in general (Conditions 1–3).

6.3. Design Implications and Future Work

6.3.1. Augmenting Existing Media Content. What this study firstly begins to illustrate is that augmenting the consumption of existing traditional 16:9 content, as demonstrated through the virtual cinema environment employed by Condition 4, can increase immersion, engagement and enjoyment. This is an important point to note as it lends relevance to consuming said content in VR as opposed to the TV, regardless of the capability for telepresence. Whilst previous augmentations of TV content have been suggested to improve immersion (e.g. smart wallpaper [Campbell et al. 2014]), HMDs offer the possibility not only of augmenting the aesthetics of the environment in which the content is being consumed, but also of consuming the content in an entirely different environment. Thus further consideration should be given to the potential augmenta-

tions made possible by VR HMDs that would improve the user's experience of existing media content.

6.3.2. Communicating Social Presence In Shared Immersive Environments. This study also established that embodied VR telepresence, and the level of immersion of a given media experience, both contribute to a more engaging, enjoyable and emotionally affective synchronous at-a-distance experience. However, this form of fully-embodied telepresence is not yet within reach of consumers, with more abstracted portrayals of remote users (such as discussed in the literature review) likely to persist until such sensing (e.g. Kinect or Playstation Camera) becomes commonplace alongside VR HMDs. Consequently, understanding what social cues can be captured and conveyed abstractly, and which cues are most meaningful, may bridge the gap between the abstract avatar and the depth camera-captured portrayal of the remote user. For example, by knowing you are attending to the same area as your partner (e.g. by conveying the action point as discussed in [Benford et al. 1995], by facilitating mutual orientation [Hindmarsh et al. 1998], or perhaps by physically actuating orientation [Gugenheimer et al. 2016]) togetherness may be fostered through a shared focus of attention, and prior concerns regarding missing events perhaps alleviated further still. This overlaps with the issue of occlusion: in wearing VR HMDs, our ability to express ourselves to others is diminished, with facial expressions and eye gaze in particular curtailed. Finding ways to capture and convey the VR HMD user's engagement, emotional investment (e.g. through facial expressions [Li et al. 2015]), and physical attention (such as in the case of ImmerseBoard [Higuchi et al. 2015] where gaze, gesture direction and intention are all conveyed to a remote partner) might help to reinforce togetherness in such experiences. Such technologies may even allow avatar-based portrayals to go beyond what is possible to be conveyed in the telepresence approach utilized here.

6.3.3. Neglect of Others: Societal Implications Of Adopting Embodied Telepresence. In addition, there are potentially negative connotations to the adoption of mixed reality telepresence. [Madary and Metzinger 2016] discuss this in terms of "neglect of others", asking:

> "What, if anything, is lost in cases of social interactions that are mediated using advanced telepresence in VR? If such losses were unnoticed, what negative effects for the human self-model could be expected?"

As [Turkle 2011] notes (from [Madary and Metzinger 2016]):

> "When these kinds of technology 'keep grandparents from making several-thousand-mile treks to see their grandchildren in person (and there is already evidence that they do), children will be denied something precious: the starchy feel of a grandmother's apron, the smell of her perfume up close, and the taste of her cooking'"

Madary and Metzinger added that:

> "Advances in technology could conceivably address Turkle's point about other perceptual modalities, but there remains a question about what may be lost even if we can create virtual content for other sense modalities... the concern remains that heavy use of such technology will lead to neglect or even animosity toward one's actual physical and social environment"

Further research will be required in order to ascertain firstly what, specifically, is not communicated or conveyed (e.g. be they subtle body cues, other sensory cues such as smell etc.) that might have a significant impact on the experience of connectedness and presence. And secondly, regarding the long term effects of having such a technology

available. For example, will couples that are geographically separated be less likely to go to the effort to meet given said technology? In which case, arguably a relationship may be harmed in the long term.

6.3.4. Mitigating Against The Physical Workload Of Immersive VR. Over the 8 minutes of viewing, the average effort expended in viewing in Condition 5 was over double that of Conditions 3 and 4. Firstly, from a purely physical standpoint, this increase in effort in 360° experiences is likely less sustainable over long periods than the effort expended in fixed-focus experiences (e.g. the TV and Cinema experiences of Conditions 3 and 4). Thus, for regular and lengthy at-a-distance communication and consumption, 360° experiences may prove impractical. Physical workload could present a significant barrier to consuming this media in durations approaching typical TV and film content. There is a counter-argument to this point: that users will self-regulate their amount of effort and manage their physical workload appropriately. However, if left to do so, their enjoyment of the 360° experience will likely be negatively impacted. Feedback from participants suggested that VR media requires a high level of engagement, as to not engage with it in this way would lead to continual feelings of missing activity occurring out of view. Indeed, perception of Condition 5 was negatively affected for a small subset of users, given the interview responses regarding fears of missing content.

Given this, to make 360° content suitable for long-term, longer-duration synchronous co-consumption, on the basis of these results we would suggest that research needs to be conducted regarding the maximum sustainable viewing effort, and in gaining an awareness of what others are attending to in the VR experience. Regarding maximum sustainable viewing effort, this is analogous to terminology used in exercise (e.g. maximum sustainable heart rate), where there is an understanding that there are physical limitations which must be managed in order to prevent exhaustion. By understanding the physical limitations of the VR HMD viewer, viewing effort could be regulated. This could be facilitated in a number of ways. For example the rendering, or creation, of 360° content which allows for ebbs and flows between recovery periods. Or where a single focus of attention is provided or emphasized, versus higher engagement periods where multiple activities occur around the viewer.

Viewing effort could also be regulated through additional cues for helping the user focus on an event. And designed for in terms of managing viewing virtually (e.g. using a game controller to move the view explicitly) or through physical actuation (e.g. SwiVRChair [Gugenheimer et al. 2016]) thus reducing physical effort. Such efforts will likely also have importance regarding solo consumption of 360° content, given early indications that the amount of effort involved may be problematic[23]. Increasing connectedness with others in the experience is also potentially worthwhile with respect to reducing fatigue. For example, the conveyance of the gaze fixation of remote users may reassure others that they are not missing out on other parts of the experience, by encouraging mutual attention to certain aspects of the VR experience. Future research might quantify the amount of effort expended longitudinally as ownership of VR HMDs becomes mainstream, and the extent to which fatigue affects adoption and usage, and would be aided in doing so by this paper providing a baseline and novel analyses to rely upon.

[23]e.g. the experience of early attempts at VR films at the Sundance Film Festival '16 , where one reviewer suggested that "damning issues... were common across the films: the need to make me, the viewer, take on the additional burdens of director and cinematographer", see http://arstechnica.co.uk/the-multiverse/2016/02/sundances-vr-films-fail-by-passing-the-workload-buck-to-their-viewers/

7. GENERAL DISCUSSION: IMPLICATIONS FOR SYNCHRONOUS AT-A-DISTANCE MEDIA CONSUMPTION IN THE HOME

Considering TV content alone, Study 1 could be seen as re-affirming the merits of TV at-a-distance, building upon an extensive body of research (predominantly observed field and laboratory studies) in this domain. However, the studies in this paper also emphasize that the design space for at-a-distance media consumption goes beyond traditional TV content. Study 1 tentatively demonstrated the utility of music at-a-distance as a shared backdrop to other activities, whilst Study 2 began to explore how immersive VR HMDs could introduce new ways to experience media together at-a-distance.

Affordable consumer VR HMDs are imminently available, and many homes already contain the necessary tools to allow for VR HMD-use at-a-distance, in the form of powerful games consoles and RGBD sensing. Consequently, Study 2 examined user preferences regarding VR HMD use for media consumption at-a-distance compared to the TV. Despite technical limitations, such as the limited field-of-view of existing consumer VR HMDs, participants significantly preferred the embodied VR telepresence (i.e. the ability to share a space with a remote correspondent) as a means of communicating, compared to picture-in-picture co-viewing. Moreover, consuming traditional TV content in an environment augmented to enhance immersion (a virtual cinema), and consuming content designed for 360° immersion in VR, both led to significant improvements regarding participant's media immersion, engagement and enjoyment in a shared experience. There exists significant research still to be done in order to facilitate such experiences (see subsection 6.3). However, such displays have the capability to support greater immersion than the TV, and enable communication that is tending towards the recreation of collocated experiences at-a-distance, the scenario that TV at-a-distance technology have previously attempted to emulate.

The potential adoption of such media types and technologies for at-a-distance use remains largely unknown. However, it will inevitably be influenced by the adoption for solus consumption. For example, 360° content represents a burgeoning domain gaining significant backing from content makers (from Disney[24] to David Attenborough[25]), and consequently may provide significant motivation for the adoption of VR HMDs in the home, acting as a gateway to mixed reality at-a-distance experiences. The findings of these studies suggest a wider design space for at-a-distance media consumption, with music, TV content (both consumed via the TV and augmented for immersion) and VR content all possibilities. This begs the question: what are the differences between these media types, and what further research will be required to understand how best to support at-a-distance media consumption in the home?

7.1. The Role Of TV And Music In A VR World: Supporting Varying Attentional Demand And Media Immersion At-A-Distance

Media consumption in this paper has loosely spanned a spectrum of media immersion and attentional demand across both media types and displays. VR HMDs diminish the users' awareness of reality to better facilitate presence in VR, and consequently media immersion. Study 2 demonstrated that this immersion improved users' enjoyment and engagement in synchronous experiences. However, Study 1 suggested that media immersion, and a high attentional demand regarding synchronous activity, is not always preferable. In the case of two couples, music served as a shared backdrop whilst attention was devoted to other, local, tasks and interactions. More broadly, the adoption of media multi-tasking behaviours (as hinted at in Study 1 by the frequent closing of the

[24]Disney Movies VR store.steampowered.com/app/469650/
[25]theguardian.com/media/2015/jul/05/david-attenborough-vr-atlantic-national-history-museum

Text chat application, and as already explored in literature regarding multi-screening ([McGill et al. 2015b; Rooksby et al. 2015]) suggests that attention to the TV and its content also varies, albeit to an unknown degree in at-a-distance contexts.

What these findings suggest is that, for at-a-distance media consumption, different media types may have the capacity to fulfil subtly different roles. This can be ascribed to their varying attentional demand and ability to co-exist with other activities. For those for whom it was acceptable to consume, music offered what appeared to be a low-attention means of engaging with a partner whilst allowing for multi-tasking and mobility. Content consumed through the TV required more attention, predominantly being a visual, sit-down activity. However, the viewer was still free to vary their attention, for example on the basis of multi-screening activity or the presence of those in the local environment (e.g. others in the living room). In contrast, VR content implicitly requires the greatest attention by occluding reality, and provides little support for multi-tasking (although in both cases this is dependent on the design of the VR experience [McGill et al. 2015a; Boland and McGill 2015] and should not be considered to be an ever-present trait of such displays).

7.2. Factors Influencing The Adoption And Usage Of At-A-Distance Media Consumption

On this basis, we would argue that at-a-distance media consumption across a range of media types (e.g. Music, TV content, IVEs) and mediums (e.g. TV, VR HMD) merits further investigation. These media types and technologies have varying attentional demands, and thus may fill different evolutionary niches in the home. Music could offer a shared backdrop to other potentially mobile activities. TV could offer a shared and engaging experience whilst still allowing for prevalent multi-screening. And VR could allow for highly immersive shared experiences to which attention is entirely devoted. However, there are likely to be a range of other significant factors which could influence adoption and the choice for consumption in any given context e.g.:

Media Content. The attentional demand of the media being consumed (e.g. music, TV, IVEs); the accessibility of content providers given variability in subscription service use and the content available across different geographic regions; and how media selections and interactions are supported.

Consumption Medium. The availability of consumption mediums (e.g. sound space, TV, mobile devices, VR/AR HMD); the capability of a given medium to present the shared experience (e.g. viewing a partner's VR HMD content on your TV or mobile device); and the capability of a given consumption medium to support multi-tasking and mobility.

Engagement. The users' engagement toward a shared experience versus other activities (i.e. what else are they doing, how invested do they wish to be and for how long do they wish to engage in a shared experience).

Relationship. The users' relationship with those they are consuming with at-a-distance (e.g. intimacy, attachment, distance and time separated).

Communication. How communication is to be facilitated and social presence conveyed (e.g. text, audio, video, telepresence), influenced by the affective benefits and costs of the communication technology being utilized; existing preferences and habits; the perceived acceptability of a given communication modality; what sensing and rendering technologies are available to facilitate said communication; and to what extent the communication modality is appropriate given the media content being consumed (e.g. speech during dialogue-heavy content).

Scale and Social Context. How many people will be sharing said experience, across how many places and at what distances. To what extent can and will other collocated persons take part, and how socially acceptable would consumption be?

Experience. The experiential qualities of the media and consumption medium (e.g. enjoyability, engagement, media immersion, fidelity) and the suitability of the media with respect to co-consumption (e.g. in terms of perceived emotional connectedness and togetherness)

Effort. What is the effort required to find and join the at-a-distance experience; how discoverable are existing experiences; how much effort is involved regarding the setup and viewing (e.g. wearing VR HMDs, using multiple devices, employing external cameras or capture equipment etc.)

Consequently, further research will be required to understand the influence these factors have on adoption and usage, and thus what types of media have a role to play in consumption at-a-distance. This would require both larger scale deployments of systems such as CastAway, as well as the implementation and / or instrumentation (of usage and communications as in Study 1) of other means of shared media consumption at-a-distance in the home (e.g. through shared sound spaces or VR HMD use). Some capacity for gauging attention and noting multi-tasking behaviour will also be necessary to establish how attention varies during at-a-distance media consumption e.g. through qualitative analysis or self-reporting or quantitative measures such as examining device activity [Rooksby et al. 2016] or gaze [Brown et al. 2014].

7.3. Supporting Users In Controlling And Conveying Attention

There is also the question as to how the underlying media consumption technology should aim to support shared experiences between users across varying media types, and varying levels of attention to / engagement with the media being consumed. This support could be in aiding transfers between consumption mediums (from a shared sound space, to TV, to mobile device, to VR HMD etc.), or in supporting the consumption of media across these boundaries (e.g. viewing a partner at-a-distance's VR HMD activity on a TV instead of having to fully immerse yourself in VR to the exclusion of other activities). Further understanding user expectations regarding the attention of others will also be required. How, and when, should attention be conveyed e.g. knowing when a partner is viewing the TV versus when they have left the room, or understanding where a partner is attending to in VR?

7.4. Supporting Communication During Consumption At-A-Distance

Understanding how users wish to communicate during these shared experiences, and supporting plasticity in communication modality use, also appears important, given the observed escalation/de-escalation behaviour in Study 1. How should users be supported in transitioning between communication modalities, and when are such transitions made? And to what extent is the usage of a given communication modality coupled to a specific media type or activity? Whilst Study 1 provided some tentative insights into this for TV and Music, further research would be required to establish whether these insights extent to the broader population, and where telepresence fits in to existing contexts (as for example made possible by AR HMDs such as Hololens or projection-based approaches such as Room2Room as discussed previously). Supporting embodied telepresence during existing media consumption might well be preferable in some contexts. But understanding what these contexts are, the extent to which this presence should be bi-directional, how transitions to and from this communication modality are managed and how others in the room are included will require further research.

Conversely, Study 2 is an unexplored example of supporting transitions in communication modalities, having only facilitated embodied telepresence in VR. Given the capacity for textual chat, less intrusive forms of communication (e.g. text or audio

only), or communication tailored toward awareness (e.g. always being able to see your partners facial expressions) variations such as those observed in Study 1 might once again emerge, and lend further insight into how best to support communication in these contexts.

7.5. Summary

From music, to TV content, to VR, media can vary in its immersive properties, the effect it has on those undertaking a shared experience and the attention and effort it demands. From textual chat, to audio, to video, to embodied telepresence, our capability to communicate and be aware of those at-a-distance is also changing. We have discussed the initial adoption of different forms of media for at-a-distance consumption (TV and music content), how communication varied during consumption (escalations to VMC for media selections) and user preferences regarding immersion and communication at-a-distance (through VR HMDs supporting telepresence). Through this, this paper serves as an initial exploration of the expanding design space of at-a-distance media consumption for the home.

8. CONCLUSIONS

This paper has firstly presented findings from an in-the-wild evaluation of a synchronous shared at-a-distance smart TV system, *CastAway*, built on an existing smart TV platform (Google Chromecast). Across five couples, for one week each, we gained initial insights into the early adoption and usage of at-a-distance media, how couples communicated, to what extent they consumed TV and music content, and the perceived benefits this system had on communications and togetherness. Secondly, in a laboratory study, this paper investigated how the impending availability of affordable consumer VR HMDs, capable of supporting embodied telepresence and increased media immersion compared to the TV, might impact media consumption at-a-distance, finding user preferences for both telepresence-based communications, consuming existing content in an immersive virtual environment (a cinema setting) and consuming immersive VR content. Finally, we discussed the implications these studies had for the near-future of consumer synchronous at-a-distance media consumption. Combined, these studies begin to explore the design space around how at-a-distance media consumption can be supported and experienced, what factors might influence usage and adoption and the implications for supporting communication and telepresence during media consumption.

ACKNOWLEDGMENTS

This work was supported in part by Bang & Olufsen. This publication only reflects the authors' views.

References

Almond Pilar N Aguila. 2011. Living long-distance relationships through computer-mediated communication. *Social Science Diliman* 5, 1&2 (2011). http://journals.upd.edu.ph/index.php/socialsciencediliman/article/viewArticle/2045

David Attenborough. 2006. Planet Earth. (2006). http://www.bbc.co.uk/programmes/b006mywy

S. Beck, A. Kunert, A. Kulik, and B. Froehlich. 2013. Immersive Group-to-Group Telepresence. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (April 2013), 616–625. DOI:http://dx.doi.org/10.1109/TVCG.2013.33

Steve Benford, John Bowers, Lennart E. Fahlén, Chris Greenhalgh, and Dave Snowdon. 1995. User Embodiment in Collaborative Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 242–249. DOI:http://dx.doi.org/10.1145/223904.223935

Regina Bernhaupt, Marianna Obrist, Astrid Weiss, Elke Beck, and Manfred Tscheligi. 2008. Trends in the Living Room and Beyond: Results from Ethnographic Studies Using Creative and

Playful Probing. *Computers in Entertainment (CIE)* 6, 1, Article 5 (May 2008), 23 pages. DOI:http://dx.doi.org/10.1145/1350843.1350848

M. Billinghurst, A. Cheok, S. Prince, and H. Kato. 2002. Real world teleconferencing. *Computer Graphics and Applications, IEEE* 22, 6 (Nov 2002), 11–13. DOI:http://dx.doi.org/10.1109/MCG.2002.1046623

Daniel Boland and Mark McGill. 2015. Lost in the Rift: Engaging with Mixed Reality. *XRDS* 22, 1 (Nov. 2015), 40–45. DOI:http://dx.doi.org/10.1145/2810046

Andy Brown, Michael Evans, Caroline Jay, Maxine Glancy, Rhianne Jones, and Simon Harper. 2014. HCI over Multiple Screens. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. ACM, New York, NY, USA, 665–674. DOI:http://dx.doi.org/10.1145/2559206.2578869

Jed R. Brubaker, Gina Venolia, and John C. Tang. 2012. Focusing on Shared Experiences: Moving Beyond the Camera in Video Communication. In *Proceedings of the Designing Interactive Systems Conference (DIS '12)*. ACM, New York, NY, USA, 96–105. DOI:http://dx.doi.org/10.1145/2317956.2317973

Duncan P. Brumby, Helena Du Toit, Harry J. Griffin, Ana Tajadura-Jiménez, and Anna L. Cox. 2014. Working with the Television on: An Investigation into Media Multitasking. In *Proceedings of the Extended Abstracts of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI EA '14)*. ACM, New York, NY, USA, 1807–1812. DOI:http://dx.doi.org/10.1145/2559206.2581210

Tatiana Buhler, Carman Neustaedter, and Serena Hillman. 2013. How and Why Teenagers Use Video Chat. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 759–768. DOI:http://dx.doi.org/10.1145/2441776.2441861

Rick Busselle and Helena Bilandzic. 2009. Measuring Narrative Engagement. *Media Psychology* 12, 4 (2009), 321–347. DOI:http://dx.doi.org/10.1080/15213260903287259

Rosie Campbell, Richard Felton, and Charlotte Hoarse. 2014. Smart Wallpaper. In *Industry Track - Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video - TVX '14*.

James J. Cummings and Jeremy N. Bailenson. 2015. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology* (2015), 1–38. DOI:http://dx.doi.org/10.1080/15213269.2015.1015740

Marianne Dainton and Brooks Aylor. 2002. Patterns of communication channel use in the maintenance of long-distance relationships. *Communication Research Reports* 19, 2 (2002), 118–129. DOI:http://dx.doi.org/10.1080/08824090209384839

Loni Dansie. 2012. *Long-distance dating relationships among college students: The benefits and drawbacks of using technology*. Master's thesis. University of Missouri–Columbia. https://mospace.umsystem.edu/xmlui/handle/10355/15248

Niloofar Dezfuli, Mohammadreza Khalilbeigi, Max Mühlhäuser, and David Geerts. 2011. A Study on Interpersonal Relationships for Social Interactive Television. In *Proceddings of the 9th International Interactive Conference on Interactive Television (EuroITV '11)*. ACM, New York, NY, USA, 21–24. DOI:http://dx.doi.org/10.1145/2000119.2000123

Discovery. 2015. Discovery Virtual Reality (360 Videos). (2015). https://www.youtube.com/playlist?list=PLiCk2I6PXl5qm0CTvO6zXt3k33GTmcIvk

Disney. 2015. Star Wars 360. (2015). https://www.facebook.com/StarWars/videos/1030579940326940/

Pierre Dragicevic. 2015. HCI Statistics without p-values. *Research Report* (2015). DOI:http://dx.doi.org/RR-8738hal-01162238

Helena du Toit. 2012. *Working while watching TV, is it really work?* Master's thesis. London University. https://www.ucl.ac.uk/uclic/studying/taught-courses/distinction-projects/2012-theses/DuToit-2012

Nicolas Ducheneaut, Robert J. Moore, Lora Oehlberg, James D. Thornton, and Eric Nickell. 2008. Social TV: Designing for Distributed, Sociable Television Viewing. *International Journal of Human–Computer Interaction* 24, 2 (feb 2008), 136–154. DOI:http://dx.doi.org/10.1080/10447310701821426

Andy Field, Jeremy Miles, and Zoë Field. 2012. *Discovering Statistics Using R*. SAGE Publications. 992 pages. https://uk.sagepub.com/en-gb/eur/discovering-statistics-using-r/book236067

Azadeh Forghani, Gina Venolia, and Kori Inkpen. 2014. Media2Gether: Sharing Media During a Call. In *Proceedings of the 18th International Conference on Supporting Group Work (GROUP '14)*. ACM, New York, NY, USA, 142–151. DOI:http://dx.doi.org/10.1145/2660398.2660417

David Geerts, Rinze Leenheer, Dirk De Grooff, Joost Negenman, and Susanne Heijstraten. 2014. In Front of and Behind the Second Screen: Viewer and Producer Perspectives on a Companion App. In *Proceedings of the 2014 ACM International Conference on Interactive Experiences for TV and Online Video (TVX '14)*. ACM, New York, NY, USA, 95–102. DOI:http://dx.doi.org/10.1145/2602299.2602312

David Geerts, Ishan Vaishnavi, Rufael Mekuria, Oskar van Deventer, and Pablo Cesar. 2011. Are We in Sync?: Synchronization Requirements for Watching Online Video Together.. In *Proceedings of the*

SIGCHI Conference on Human Factors in Computing Systems (CHI '11). ACM, New York, NY, USA, 311–314. DOI:http://dx.doi.org/10.1145/1978942.1978986

Anand Giridharadas. 2014. For Couples Split by Distance, Two Screens Can Blink as One. (2014). http://www.nytimes.com/2014/01/10/arts/television/for-couples-split-by-distance-two-screens-can-blink-as-one.html?

Daniel Gooch and Leon Watts. 2012. YourGloves, Hothands and Hotmits: Devices to Hold Hands at a Distance. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 157–166. DOI:http://dx.doi.org/10.1145/2380116.2380138

Google. 2015. Chromecast. (2015). http://www.google.co.uk/chrome/devices/chromecast/

Chris Greenhalgh and Steven Benford. 1995. MASSIVE: A Collaborative Virtual Environment for Teleconferencing. *ACM Transactions on Computer-Human Interaction (TOCHI)* 2, 3 (Sept. 1995), 239–261. DOI:http://dx.doi.org/10.1145/210079.210088

Jan Gugenheimer, Wolf Rukzio, and Enrico Rukzio. 2016. SwiVRChair: A Motorized Swivel Chair to Nudge Users' Orientation for 360 Degree Storytelling in Virtual Reality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. https://www.uni-ulm.de/in/mi/mi-forschung/mi-forschung-rukzio/projects/swivrchair.html

Jeff Hancock, Jeremy Birnholtz, Natalya Bazarova, Jamie Guillory, Josh Perlin, and Barrett Amos. 2009. Butler Lies: Awareness, Deception and Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 517–526. DOI:http://dx.doi.org/10.1145/1518701.1518782

Gunnar Harboe, Noel Massey, Crysta Metcalf, David Wheatley, and Guy Romano. 2008a. The Uses of Social Television. *Computers in Entertainment (CIE) - Social television and user interaction* 6, 1, Article 8 (May 2008), 15 pages. DOI:http://dx.doi.org/10.1145/1350843.1350851

Gunnar Harboe, Crysta J. Metcalf, Frank Bentley, Joe Tullio, Noel Massey, and Guy Romano. 2008b. Ambient Social Tv: Drawing People into a Shared Experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 1–10. DOI:http://dx.doi.org/10.1145/1357054.1357056

Marc Hassenzahl, Stephanie Heidecker, Kai Eckoldt, Sarah Diefenbach, and Uwe Hillmann. 2012. All You Need is Love: Current Strategies of Mediating Intimate Relationships Through Technology. *ACM Transactions on Computer-Human Interaction (TOCHI)* 19, 4, Article 30 (Dec. 2012), 19 pages. DOI:http://dx.doi.org/10.1145/2395131.2395137

Jan Hess, Benedikt Ley, Corinna Ogonowski, Lin Wan, and Volker Wulf. 2011. Jumping Between Devices and Services: Towards an Integrated Concept for Social Tv. In *Proceddings of the 9th International Interactive Conference on Interactive Television (EuroITV '11)*. ACM, New York, NY, USA, 11–20. DOI:http://dx.doi.org/10.1145/2000119.2000122

Keita Higuchi, Yinpeng Chen, Philip A. Chou, Zhengyou Zhang, and Zicheng Liu. 2015. ImmerseBoard: Immersive Telepresence Experience Using a Digital Whiteboard. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2383–2392. DOI:http://dx.doi.org/10.1145/2702123.2702160

Jon Hindmarsh, Mike Fraser, Christian Heath, Steve Benford, and Chris Greenhalgh. 1998. Fragmented Interaction: Establishing Mutual Orientation in Virtual Environments. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work (CSCW '98)*. ACM, New York, NY, USA, 217–226. DOI:http://dx.doi.org/10.1145/289444.289496

Jerry L. Hintze and Ray D. Nelson. 1998. Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician* 52, 2 (1998), 181–184. DOI:http://dx.doi.org/10.1080/00031305.1998.10480559

Kori Inkpen, Brett Taylor, Sasa Junuzovic, John Tang, and Gina Venolia. 2013. Experiences2Go: Sharing Kids' Activities Outside the Home with Remote Family Members. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 1329–1340. DOI:http://dx.doi.org/10.1145/2441776.2441926

Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and Defining the Experience of Immersion in Games. *International Journal of Human-Computer Studies* 66, 9 (Sept. 2008), 641–661. DOI:http://dx.doi.org/10.1016/j.ijhcs.2008.04.004

Martin Krzywinski and Naomi Altman. 2014. Points of Significance: Visualizing samples with box plots. *Nature Methods* 11, 2 (jan 2014), 119–120. DOI:http://dx.doi.org/10.1038/nmeth.2813

Hao Li, Laura Trutoiu, Kyle Olszewski, Lingyu Wei, Tristan Trutna, Pei-Lun Hsieh, Aaron Nicholls, and Chongyang Ma. 2015. Facial Performance Sensing Head-mounted Display. *ACM Trans. Graph.* 34, 4, Article 47 (July 2015), 9 pages. DOI:http://dx.doi.org/10.1145/2766939

M Lombard, L Weinstein, and T Ditton. 2011. Measuring telepresence: the validity of the Temple Presence Inventory (TPI) in a gaming context. (2011).

Anna Macaranas, Gina Venolia, Kori Inkpen, and John Tang. 2013. *Human-Computer Interaction – INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part IV*. Springer Berlin Heidelberg, Berlin, Heidelberg, Chapter Sharing Experiences over Video: Watching Video Programs together at a Distance, 73–90. DOI:http://dx.doi.org/10.1007/978-3-642-40498-6_5

Michael Madary and Thomas K. Metzinger. 2016. Real Virtuality: A Code of Ethical Conduct Recommendations for Good Scientific Practice and the Consumers of VR-Technology. *Frontiers in Robotics and AI* 3, 3 (2016). DOI:http://dx.doi.org/10.3389/frobt.2016.00003

Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015a. A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2143–2152. DOI:http://dx.doi.org/10.1145/2702123.2702382

Mark McGill, John H. Williamson, and Stephen A. Brewster. 2015b. A review of collocated multi-user TV. *Personal and Ubiquitous Computing* 19, 5 (2015), 743–759. DOI:http://dx.doi.org/10.1007/s00779-015-0860-1

Crysta Metcalf, Gunnar Harboe, Joe Tullio, Noel Massey, Guy Romano, Elaine M. Huang, and Frank Bentley. 2008. Examining Presence and Lightweight Messaging in a Social Television Experience. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 4, 4, Article 27 (Nov. 2008), 16 pages. DOI:http://dx.doi.org/10.1145/1412196.1412200

Microsoft. 2015. HoloLens. (2015). https://www.microsoft.com/microsoft-hololens/en-us

P Milgram and H Colquhoun. 1999. A taxonomy of real and virtual world display integration. *Mixed reality: Merging real and virtual worlds* (1999), 5–30. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.32.6230

Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321—-1329. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.102.4646

Evan Narcisse. 2011. Say Goodbye to Netflix's Party Mode on the Xbox 360. (2011). http://kotaku.com/5866245/say-goodbye-to-netflixs-party-mode-on-the-xbox-360

Mukesh Nathan, Chris Harrison, Svetlana Yarosh, Loren Terveen, Larry Stead, and Brian Amento. 2008. CollaboraTV: Making Television Viewing Social Again. In *Proceedings of the 1st International Conference on Designing Interactive User Experiences for TV and Video (UXTV '08)*. ACM, New York, NY, USA, 85–94. DOI:http://dx.doi.org/10.1145/1453805.1453824

Carman Neustaedter and Saul Greenberg. 2012. Intimacy in Long-distance Relationships over Video Chat. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 753–762. DOI:http://dx.doi.org/10.1145/2207676.2207785

Oculus. 2015. Oculus Rift DK2. (2015). https://www.oculus.com/en-us/dk2/

Jarmo Palviainen, Kati Kuusinen, and Kaisa Väänänänen-Vainio-Mattila. 2013. Designing for Presence in Social Television Interaction. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia (MUM '13)*. ACM, New York, NY, USA, Article 9, 10 pages. DOI:http://dx.doi.org/10.1145/2541831.2541860

Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. 2016. Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1716–1725. DOI:http://dx.doi.org/10.1145/2818048.2819965

Renderheads. 2015. AVPro Windows Media. (2015). http://renderheads.com/product/av-pro-windows-media/

John Rooksby, Parvin Asadzadeh, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2016. Personal Tracking of Screen Time on Digital Devices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 284–296. DOI:http://dx.doi.org/10.1145/2858036.2858055

John Rooksby, Timothy E. Smith, Alistair Morrison, Mattias Rost, and Matthew Chalmers. 2015. *Configuring Attention in the Multiscreen Living Room*. Springer International Publishing, Cham, 243–261. DOI:http://dx.doi.org/10.1007/978-3-319-20499-4_13

Mattias Rost, Christos Kitsos, Alexander Morgan, Martin Podlubny, Pietro Romeo, Edoardo Russo, and Matthew Chalmers. 2016. Forget-me-not: History-less Mobile Messaging. In *Proceedings of the 34rd Annual ACM Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA. DOI:http://dx.doi.org/10.1145/2858036.2858347

Rovo89 and Tungstwenty. 2015. Xposed Framework. (2015). http://repo.xposed.info/module/de.robv.android.xposed.installer

J Saldaña. 2015. *The coding manual for qualitative researchers*. https://uk.sagepub.com/en-gb/eur/the-coding-manual-for-qualitative-researchers/book243616

Steven Schirra, Huan Sun, and Frank Bentley. 2014. Together Alone: Motivations for Live-tweeting a Television Series. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2441–2450. DOI:http://dx.doi.org/10.1145/2556288.2557070

David A. Shamma and Yiming Liu. 2009. Zync with me: Synchronized sharing of video through instant messaging. *Social Interactive Television: Immersive Shared Experiences and Perspectives* (2009). http://www.igi-global.com/chapter/social-interactive-television/29211

Dong-Hee Shin. 2013. Defining sociability and social presence in Social TV. *Computers in human behavior* 29, 3 (2013), 939–947. http://dx.doi.org/10.1016/j.chb.2012.07.006

Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3549–3557. DOI:http://dx.doi.org/10.1098/rstb.2009.0138

Laura Stafford. 2005. *Maintaining Long-Distance and Cross-Residential Relationships*. Routledge.

Sherry Turkle. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books, Inc., New York, NY, USA. http://dl.acm.org/citation.cfm?id=1972496

Radu-Daniel Vatavu. 2015. Audience Silhouettes: Peripheral Awareness of Synchronous Audience Kinesics for Social Television. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video (TVX '15)*. ACM, New York, NY, USA, 13–22. DOI:http://dx.doi.org/10.1145/2745197.2745207

Justin D. Weisz, Sara Kiesler, Hui Zhang, Yuqing Ren, Robert E. Kraut, and Joseph A. Konstan. 2007. Watching Together: Integrating Text Chat with Video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 877–886. DOI:http://dx.doi.org/10.1145/1240624.1240756

Doug Williams, Marian F. Ursu, Pablo Cesar, Karl Bergström, Ian Kegel, and Joshan Meenowa. 2009. An Emergent Role for TV in Social Communication. In *Proceedings of the Seventh European Conference on European Interactive Television Conference (EuroITV '09)*. ACM, New York, NY, USA, 19–28. DOI:http://dx.doi.org/10.1145/1542084.1542088

Doug Williams, Marian F. Ursu, Joshan Meenowa, Pablo Cesar, Ian Kegel, and Karl Bergström. 2011. Video mediated social interaction between groups: System requirements and technology challenges. *Telematics and Informatics* 28, 4 (2011), 251 – 270. DOI:http://dx.doi.org/10.1016/j.tele.2010.11.001 Television in a digital era - Usage and policy issues.

Svetlana Yarosh, Panos Markopoulos, and Gregory D. Abowd. 2014. Towards a Questionnaire for Measuring Affective Benefits and Costs of Communication Technologies. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 84–96. DOI:http://dx.doi.org/10.1145/2531602.2531634

Received May 2016; revised NA; accepted NA

Online Appendix to: Examining The Role Of Smart TVs And Virtual Reality Head-Mounted Displays In Synchronous At-A-Distance Media Consumption

Mark McGill, University of Glasgow
John H. Williamson, University of Glasgow
Stephen Brewster, University of Glasgow

A. CASTAWAY QUESTIONNAIRES

See online submission for attached ZIP of questionnaires.

B. VR QUESTIONNAIRE

See online submission.

C. VIDEO

See online submission for attached MP4. For convenience, there is also a private link to the video here: https://www.youtube.com/watch?v=WPkeabY-W9A&feature=youtu.be.

D. STUDY 1 INTERVIEWS

Full unedited excerpts from the interviews for each theme can be found below.

D.1. Attitudes Toward CastAway: At-a-distance media consumption was universally liked, but can exacerbate feelings of separation

P2: I really liked it, I thought. I would use it long term, to be honest.
P1: I would... if it was on my iPhone! Because I don't like Android. (laughter). Sorry! But yeah, I would, it was good, I enjoyed it. I think I'll miss it!
P1: I enjoyed it. I think I'll miss it!

P4: Overall I really enjoyed the experience – it was fun and new and something I would never have tried.

P5: I just think it's really good, I think if it were available it'd probably be something I would use.
P6: See if you had the Chromecasts and were just using our own phones, then I would definitely use that, and we'd keep doing it.

P7: I would use the system.
P8: Yeh, with more options [TV content] it would be better.

P9: We liked it.
P10: We would use it. Last year we were both on years abroad, and we tried sometimes to do that, but have [the] Skype app [connected] and watch something at the same time. But we'd always have problems, like someone's internet wouldn't work, or it would be out of sync and... it was more hassle than it's worth. But with something like this it would have been really good.
P9: And also our laptops would be really slow, because you'd be running the TV program or whatever and also have Skype in the background, and it wouldn't really work. The fact this was all in one kind of made it a lot easier.

© 2016 Copyright held by the owner/author(s). 1073-0516/2016/02-ART0 $15.00
DOI: 0000001.0000001

P5: It kind of made you more aware of the fact that you weren't just sitting watching TV together [in person]. But it was nice, it felt that you were making time for each other, sitting down and watching something together.
I: Did it make it worse at all that you were being reminded that you weren't in the same space?
P6: At times yeh.
P5: Sometimes yeh.
P6: At times yeh... When we were watching that we're really enjoying, and having proper conversations about it, it kind of reminded you that it'd be nice if you were actually together.
P5: But I suppose you don't have any other option. It's not as good as actually sitting together watching something, but it's better than watching stuff at different times and talking about it later.
I: The benefits of the system outweighed the negatives of reminding you that you weren't in the same space?
P5: Yeh, it was a good middle ground between actually watching something together and... not!

D.2. Effects of Usage: Increased connectedness and communications

P3: I would say so, slightly more connected, just in terms of the fact that one of the things you don't get to do if you're living apart is watch things simultaneously, it takes a lot more effort to do that anyway, you need to like, whereas with this you press a button and that's it.
P4: Yes – I think it made us feel closer and more connected because we would arrange to watch something then could discuss it. Made it more fun and intriguing watching something knowing he was too and we could then discuss during or after.

P6: It helped.
P5: It was good, it was nice, the idea that you'd make time for each other, we'd watch a lot of the same stuff anyway, it was nice to actually sit down and watch it together and talk about it as you were watching it.
P6: It helped... We have a lot of communication anyway, so it's not like it added more communication to our relationship, it was more the point that we were actually doing it, watching something, together.

P7: Help I'd say.
P8: Yeh, it was good to watch stuff together.

P9: It helped.
P10: I'd agree.
P9: It helped... Definitely didn't hinder. It was just like a different way to talk to each other really. And do something we would have done if we were together, but we didn't have to be together.
P10: Or do something that we would have done separately, but be able to share it.
I: Did the system make you feel closer or more connected?
P9: Closer.
P10: I'd say so, I guess just by sharing an experience.
P9: Closer... It's something you could have done anyway if you synced it up beforehand.
P10: But it makes it a lot easier.

P9: A lot easier.... the little things, like making sure it's synced by just pressing the button, even though the other persons actually watching it.

D.3. Attitudes Toward TV: TV at-a-distance was universally liked

I: Did you find that having the TV made you feel closer to your partner?
P1: We watched TV and then he was like "aww this is happening" but we were in sync, so I could see what was happening at the same time. So that was good, because we could talk to each other whilst something was going on on TV.
P2: Watch Wallace and Gromit!
P1: Yeah we did (laughter).

P3: That was something that I would definitely see both of us using, because we've tried to do that in the past without an app to do it. It's definitely good to have that thing for that purpose, and having that would probably help you do it more often.
P4: I really liked the synchronised TV - we watch a lot of programmes together and often I need to wait until we are together to do that. I feel this solved it. I would definitely use this system a lot.

P5: We talk about TV, but we'd talk about it after as we'd watch it at different times.
P6: But it's easier to sit and talk about it if you're watching it at the same time, like if he was to laugh at something I'd know exactly what he's laughing at, rather than a whole paragraph to explain, because we were watching at different times.

P7: I did like that, I thought it was good. *P7*: Yeh.
P8: Because we ended up having to watch things we wouldn't normally watch.
P7: But it was good, yeh.

P9: For that I think it was perfect. It's things we would do anyway, we would both watch a tv program, maybe not necessarily at the same time, but to do it at the same time with someone else.
P10: And then to be able to have like messaging and stuff where you can talk about it.
P9: Or see the other persons face while they were watching it. Yeh, I think it's a really good idea.

P5: It'd be good if we could have had Netflix or something as well obviously.

P8: Only if it was more than BBC iPlayer, because we ended up having to watch things we wouldn't normally watch

P1: He picked the programs for TV, because he's fussy.

P3: We basically picked on different nights, who would watch what.

P5: [We would] see who wants to watch something. There were times where we knew we'd be able to watch TV together where we both looked on iPlayer to see if there was anything we fancied watching.
P6: It was more planning what time to do it, and then finding something

then. I was going to suggest something, then the next night you suggested what to put on.

P8: We spoke over the text about what to put on really.

P10: I'd say with the TV we just looked at what the options were
P9: Yeh, we discussed what we wanted to watch while we were choosing.

D.4. Attitudes Toward Music: Adoption of music at-a-distance was dichotomous

I: Did you have a preference between the TV and the music, in terms of what you enjoyed consuming together more?
P2: Probably music
P1: (At same time) Music probably, yeah, because we could be doing other things and then music would be playing and we could still talk to each other, whereas TV, like, I have very different tastes in TV than he does, and because it was BBC iPlayer, but we found a couple of things, so that was ok. I think if you had Netflix or something, that would be good, so if you could watch a full movie.
P2: I quite liked watching TV, but... the music was really good, because you were showing me songs that I'd never heard before, like songs you'd been listening to on Spotify, because you use Spotify and I don't, so like songs that [my partner] had heard and said "you'll like this", and then we were listening to songs we had listened to years ago and hadn't heard in ages, just for a laugh and stuff, cheesy songs.

P7: That [Music] was good, because normally if we're sitting studying in our own houses, we're not really connected. But allowing us to listen to the same music at the same time helped make us feel closer.
P8: And speaking about the same kind of music.
I: Did your tastes in music differ?
P8: He put Westlife on for me for a wee bit!
P7: There's music I like and music you like, and there's music we both like, so we just stuck that on.
P8: We know what bands we both like.

P3: I liked it because it's always good to be able to listen to music with someone else, but it's not the same as sitting down and putting on a CD with the person with you. It has advantages but it wasn't the kind of thing that I'd plan to do, I don't think I'd ever say lets put on this album and listen to it over the system, we'd meet up and do that. And I don't think it's something we would spontaneously do, you'd probably send a link to the video for the song instead. I liked the fact you can do that, but I don't know if I'd use it. When it's Music you want to have a proper conversation about it, which is a bit more difficult over a longer distance.
P4: I don't think the synchronised music was very good - I don't see a time I would ever use it. Maybe if we were both getting ready to go out but even then I think we would just play our own music. It was fun to pick songs that one time but I don't think I'd use it again.

P6: Disliked it.
P5: Yeh, the only times we'd listen to music together would be when we were going out.
P6: Our music taste differs a bit from our TV taste, so when we did do it, it was songs we both liked, but we weren't totally into it because our music

differs.
P5: Most of the music I would listen to is not the kind of music you'd sit down and listen to together. *I*: Would you say your tastes in TV are closer in that case?
P6: Yeh

P10: I didn't like the music as much, but I feel like we probably wouldn't use it.
P9: We don't really listen to music together that often.
P10: I guess it depends on the person. Like I think music is quite a personal thing.
P9: I would agree with that, it's maybe the way our relationship is.
I: Would you say you have similar or divergent tastes in music?
P9: Similar. I don't think it's to do with the tastes, I think it's just that I don't listen to music with other people anyway, it's a thing I do to relax.
P10: And then I guess that's not necessarily at the same time that I would want to listen.
P9: Television or films made more sense. It's more of a community based thing.

D.5. The Role of Music: A low-attention, mobile background activity

P1: We could be doing other things and then music would be playing and we could still talk to each other. I came home from work and I was just lying in bed being lazy, and he just put songs up and I was like "yeah, this is good", and I'd just sit there, not doing anything, just listening... we could be doing other things and then music would be playing and we could still talk to each other.

P8: You could walk away from it and come back in and talk about it again. Music is more... walking about.
P7: Yeh, you could just have it on in the background almost.
I: Is that what you used the music for?
P7: Music was more when we were doing stuff.
P8: I was tidying my room and stuff, and I could just hear the music playing through the TV.
P7: I think that's the nice thing, that's why it's good. With the music it brings you closer.

D.6. The Role of TV: A sit-down activity demanding attention

P2: We sat down to watch it
P1: We both sat down to watch it.
I: So you didn't have it just playing in the background?
P1: No, we just sat down to watch it, and spoke to each other while it was on. Depends what it is though, if it's something I'm interested in, I will watch it, but if it's something I'm not interested in, I'll tell him I'm watching it, but I won't watch it! (laughter)
P2: The other one I think I was just kinda watching and you really weren't.
P1: Was that the comedy one? I don't like comedy. But I was watching it, but I was also talking to you at the same time, I wasn't doing anything else.
P2: You kept talking to me, and I was kinda ignoring you a wee bit, because I was watching it. (laughter)

D.7. Attitudes Toward Communication

I: What did you prefer out of the different ways you had to chat to each other?
P1: The audio, and the video.

P3: Yeh, the [Text] messaging was the one we used the most, just because the video was quite small and again if you're watching something, you don't want to have a video up of someone else most of the time. And if you were going to video someone you'd use a different application to do that.

P6: Text
P5: We used Text, we tried them all, but we didn't use any except text, when you're trying to watch something [it was better].
I: Was there any reason you preferred text?
P5: We don't really chat on the phone that much, because we're both quite busy. Especially if you're watching something.
P6: Just using the text option fitted with us, and what we do. I think if we went out of our way to do phone calls, we wouldn't have enjoyed it as much.
P5: We used Text, we tried them all, but we didn't use any except text, when you're trying to watch something [it was better]... We don't really chat on the phone that much, because we're both quite busy. Especially if you're watching something.
P6: Just using the text option fitted with us, and what we do. I think if we went out of our way to do phone calls, we wouldn't have enjoyed it as much.

P7: It was easier, we were more constantly texting, but we were sometimes talking over video depending on what we were doing.

P9: [We used] Text. I don't think we used audio at all.
P10: Text mainly.
P9: I don't think we used audio at all.
P10: Only a little bit at the start.
P9: To discuss what we wanted to talk about. But when you're watching a TV program you've already got a visual element and a vocal element, so the text was perfect.

P2: They were quite good, because you could make [the video] as big or small as you want.
P1: For just talking to each other, it was a lot easier than using Skype or any of those things.
P2: Aye, it was so much quicker to just connect everything.
P1: And you could have the audio, or the video, I could leave the phone on my bed and walk about the room while he's still talking to me, and do other things.
P2: I never really got into Skype or anything like that.
P1: We tried Facebook video and we didn't like it either.
P2: I don't know why we didn't like it, because in theory they are all the same, but it just kinda felt right. It was good to be able to see each other

and talk, and have a background. You know you can minimize the chat, so you can do something in the background.

P3: If you're watching something, you don't want to have a video up of someone else most of the time.

P8: Using the wee video thing was good, it was better than Facetime where you can't do anything else.
P7: The voice thing was good as well because you could quickly go into a message, as opposed to phoning.
P7: The three options [overlayed Text/Audio/Video], I think that's a really good idea, I like that. Having it overlayed is helpful obviously.
P8: It's so much quicker than the way Facetime and everything is, phoning. And also I could just say something instead of waiting for you to answer the phone, I could just say what I wanted to say.

P9: When you're watching a TV program you've already got a visual element and a vocal element, so the text was perfect.

D.8. Content Synchronization: Synchronized content and control allow implicit communication of availability

P2: The sync was actually quite good, because [my partner] was behind me and I was saying "this is coming up", and then I re-synced it and we were quite good.

P5: Sometimes we'd be watching something for about an hour, and it was like 9 seconds out roughly, but we didn't really bother [re-synchronizing] because it was close enough that you were watching something. Whereas if you were watching something normally, and one person pauses it for a break then you'd be 10 minutes out! It was good we could both pause it.

P7: Almost knowing it was synchronous is better.
P8: And pausing it when we walk away, like sometimes I pause stuff and leave the room and then he's in front of me and I'm behind, it was better that we were together.

P9: The little things [made it easier], like making sure it's synced by just pressing the button, even though the other persons actually watching it.