European Commission

HORIZON 2020

The EU Framework Programme for Research and Innovation

Directorate General for Communications Networks, Content and Technology
Innovation Action

ICT-687655

2IMMERSE

D3.1 General Concepts and Challenges for Customizable Multi-Screen Interactions

Due date of deliverable: 30 June 2016

Actual submission date: 30 August 2016

Resubmitted: 28 April 2017

Start date of project: 1 December 2015 Duration: 36 months

Lead contractor for this deliverable: **BBC**

Version: **28 April 2017**

Confidentiality status: **Public**

Abstract

This document outlines the concepts at play in the design of the multi-screen experiences under development in 2-IMMERSE for the project's four pilots. Here we outline the types of experiments and the decisions they are intended to inform within and across the four scenarios of drama and sport in private and public venues.

Target audience

This is a public deliverable and could be read by anyone with an interest in the details of the experience designs being developed by the 2-IMMERSE project. We assume the audience is familiar with television and Internet technologies. This document will be read by the Project Consortium as it implements the trials to be delivered during the project.

Disclaimer

This document contains material, which is the copyright of certain 2-IMMERSE consortium parties, and may not be reproduced or copied without permission. All 2-IMMERSE consortium parties have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the 2-IMMERSE consortium as a whole, nor a certain party of the 2-IMMERSE consortium warrant that the information contained in this document is capable of use, or that use of the information is free from risk, and accept no liability for loss or damage suffered by any person using this information.

This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content.

Impressum

Full project title: 2-IMMERSE

Title of the workpackage: WP2 Distributed Media Application Platform

Document title: D3.1 General Concepts and Ideas for Multi-Screen Interactions

Editor: Phil Stenton, BBC

Workpackage Leader: Maxine Glancy, BBC

Project Co-ordinator: Helene Waters, BBC

Project Leader: Phil Stenton, BBC

This project is co-funded by the European Union through the ICT programme under Horizon2020

Copyright notice

1. Executive Summary

The 2-IMMERSE project proposes to create an extensible platform to deliver new media services that deliver flexibly configured multi-screen experiences. Unlike previous projects delivering multi-screen content, the 2-IMMERSE platform will support the orchestration of media across devices by stakeholders along the broadcast value chain. This will allow producers, broadcasters, venue owners and audiences to select content micro-services we call Distributed Media Applications (DMApps) and determine their distribution within and across devices. The challenge can be summarized as co-producing the preferred content for the preferred screen at the right time whilst allowing a satisfying mix of control across the value chain, this being in extremis between the curation of the producer and the level of control desired by the user.

In addition to the challenges of synchronization and coherence, this ambition raises new challenges for layout and control and the distribution and composition of audio, video and data further down the distribution channel, sometimes in the consumer devices of the audiences.

The extensible 2-IMMERSE platform will be developed and tested through the delivery of four service pilots: Theatre in the home and school and sport in the home (MotoGP) and in a public venue (Football). Trials of these services will use high value content from the Royal Shakespeare Company broadcasts, MotoGP content licensed by Dorna and the FA Cup final licensed by the Football Association to the BBC and BT.

This document is the first in a series of WP3 deliverables that describe design and user research challenges involved in developing these customizable, multi-screen audience experiences for home and public venues. The goal of this report is a brief introduction the general concepts and challenges for the platform. Future deliverables in this series will describe in more detail the specific design processes and experiments that lead to each service pilot specification. In the WP4 deliverables the full details of the trial of each pilot service will be described.

In this document Lab tests of audio technology for video-chat across homes and user experiments on synchronization and latency tolerance are described as examples of WP3 activities to come, that will inform the design of general capabilities that will be reused across service pilots. Design activities that lead to more general platform capabilities such as ‘on-boarding’ and capabilities specific to individual pilots will be covered as they arise in future deliverables in this series.

List of Authors

Phil Stenton - BBC
Mark Lomas - BBC
Vinoba Vinayagmoorthy - BBC
Doug Willians – BT

Reviewer

Maxine Glancy - BBC

Table of contents

2. Introduction

An extensible 2-IMMERSE platform will be developed and tested through the development and user testing of 4 service pilots: Theatre in the home and school and sport in the home (MotoGP) and in a public venue (Football). Trials of these services will use high value content from the Royal Shakespeare Company broadcasts, MotoGP content licensed by Dorna and the FA Cup final licensed by the Football Association to the BBC and BT.

This document is the first in a series WP3 deliverables that describe design and user research challenges involved in developing these customizable, multi-screen audience experiences for home and public venues. The goal of this report is a brief introduction the general concepts and challenges for the platform and future deliverables in this series will describe in more detail the specific design processes and experiments that lead to each service pilot specification.

Despite a wealth of devices, a huge choice of digital content and increasing access to high-speed broadband, today's TV experiences largely remain single screen experiences whether delivered through traditional broadcast or via broadband. Applications on 'second' devices have achieved only limited success as independent 'companions' to broadcast TV experiences. Surveys show that, although around 80% of people are using a second device (phone, tablet or laptop) when watching TV only 20% of them are engaging with 'companion content' – content that is created to accompany the broadcast content. The challenge for broadcasters is to engage audiences in a world where connected devices offer greater competition for attention and immersion.

Interface specifications will be developed collaboratively, through regular conference calls, face-to-face meetings across the team and consultation with communities of practice such as the Royal Shakespeare Company, schools and sports broadcasters.

Academic and industry research has explored more deeply the design and delivery of broadcast and broadband experiences. Early work including that of the partners in this project [1][2][3][4] has examined social network activity and TV viewing. Cesar et al [1] describe 'sharing' as one of the core grounding features of second screen experiences. Basapur et al [2][3] evaluated broadcast and socially generated companion information describing the value of the information and a preference for continuing the 'lean back' nature of TV engagement. Lochrie and Coulton [4] in their study of tweets around the reality TV show The X-Factor concluded that live tweets about the show significantly correlated with the content of the show (from moment to moment). They suggested that mobile devices were becoming a second screen not through broadcaster intervention but viewers were themselves creating forums for discussion. With the growth of the app ecosystem, broadcasters have launched apps aimed at augmenting the TV experience [5][6][7]. Companion apps such as these encourage TV audience participation. In the case of The X-Factor [5] audiences can vote through the app for the contestants and the songs they have to sing. The BBC Antiques Roadshow app [6] allowed viewers to guess the value of antiques being shown. The Channel 4 Million Pound Drop app [7] that allows you to play along had a million downloads in its first three months. At TVX 2014 Murray et al [8] demonstrated a companion app for long form TV narratives: in this case, to keep track of the plots and characters in Game of Thrones.

The BBC and others have published research on attending to more than one screen when engaging with synchronized companion content across devices [9][10][11][12]. Holmes et al

[9] monitored time spent attending to TV and second screen. Vinayagamoorthy et al [10] uncovered a complex interaction between participants and the content on the two screens. Brown et al [11] and Kern et al [12] looked at the triggers and attention switching across the two screens.

Building on and extending the work by the BBC and IRT in the MediaScape project (EU FP7 610401) [13] we aim to create an extensible, open source platform for developing coherent multi-screen experiences that are customizable at any stage in the value chain. Across the four pilots we will mix curation and customization enabling content and data feeds to be tailored to fan or venue's requirements whilst respecting a producer's artistic direction. This may require production-quality content playback and manipulation to be developed within the media client or through cloud services. An ability to personalise an experience in response to "in-the-moment layout needs" is a concept illustrated within the Fresco lab demo at CISCO's Bedfont Lakes site.

3. Contributions and Challenges

Figure 1 shows the 5 contributions of the 2-IMMERSE project.

1. The Service Development and Delivery Platform (Infrastructure) will support the 4 service pilots and be the seed for third party development to support other services.
2. A specification and seed implementation of object-based micro-services (DMApps). Designing an experience will mean selecting a constellation of these and the timing and location of their delivery.
3. Responsive composition and orchestration (control) of content and layout across devices and locations.
4. Authentic service pilots built alongside existing broadcasts with AAA licensed content (i.e. RSC performances and internationally broadcast sports events) This will require careful relationship management to secure support for releasing raw content streams and capturing any extra feeds and data we may require.
5. Production and orchestration tools across the broadcast value chain and an understanding of the new demands on production workflows and craft.

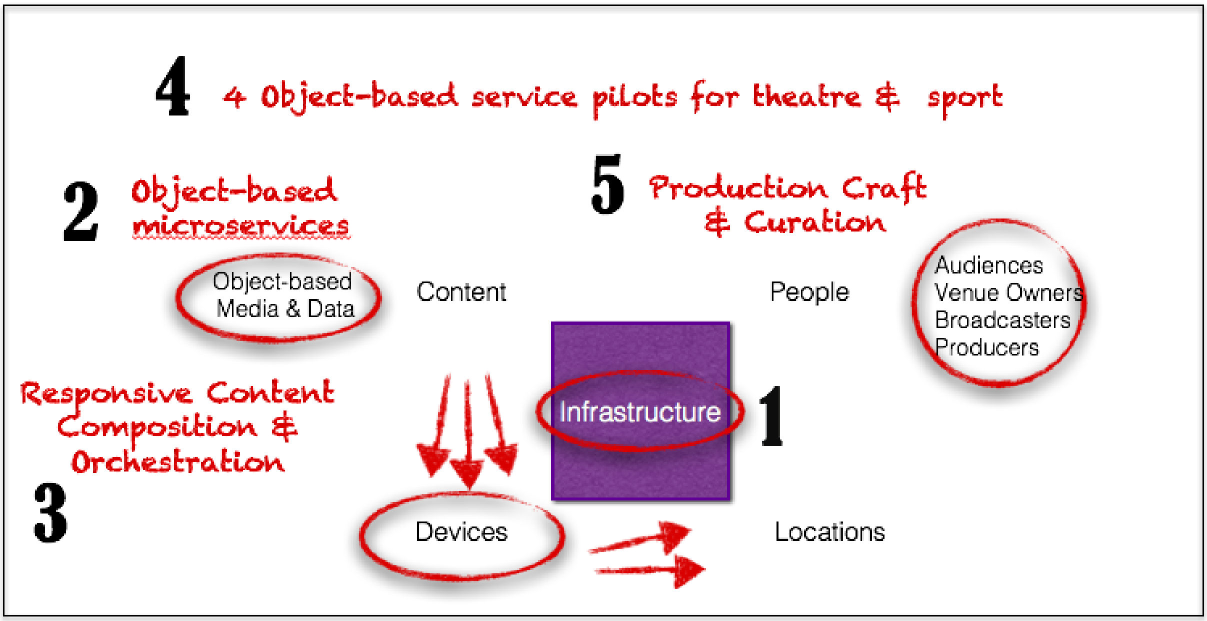

4
4 Object-based service pilots for theatre & sport
2
Object-based microservices
5
Production Craft & Curation
Object-based Media & Data
Content
People
Audiences
Venue Owners
Broadcasters
Producers
Responsive Content Composition & Orchestration
Infrastructure
1
3
Devices
Locations

Figure 1. The contributions of 2-IMMERSE

In addition to the challenges of synchronization and coherence, this ambition raises new challenges for layout and control and the distribution and composition of audio, video and data further down the distribution channel, sometimes in the consumer devices of the audiences. Figure 2 shows the challenges faced to deliver the contributions above.

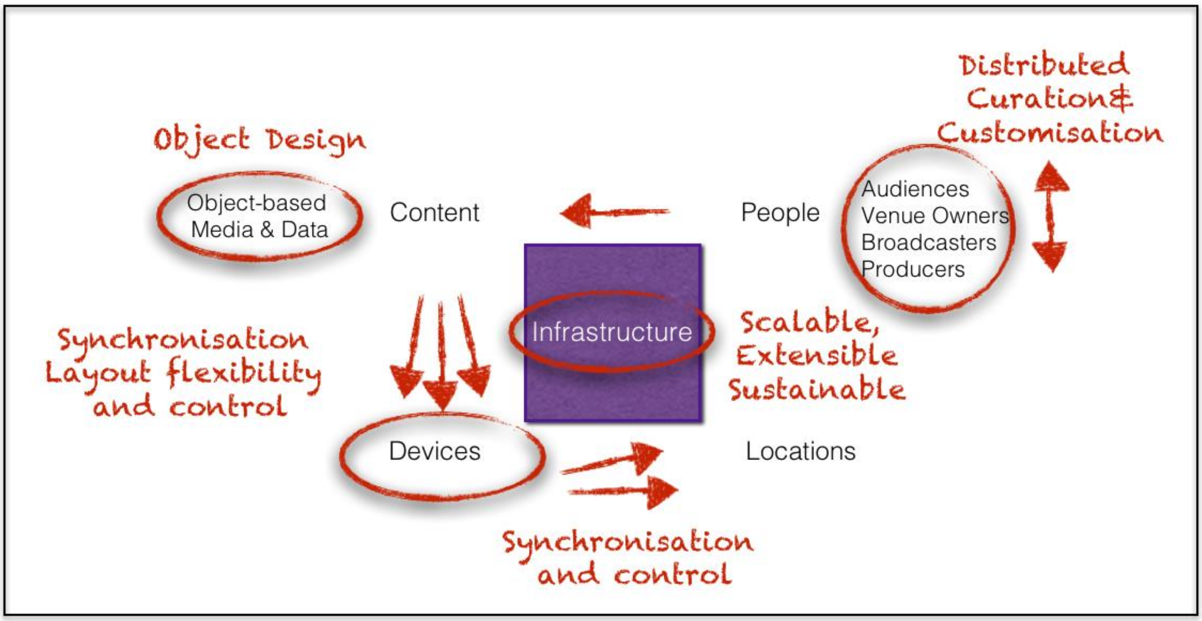

Distributed Curation& Customisation
Object Design
Object-based Media & Data
Content
People
Audiences
Venue Owners
Broadcasters
Producers
Synchronisation Layout flexibility and control
Infrastructure
Scalable, Extensible Sustainable
Devices
Locations
Synchronisation and control

Figure 2. The challenges of 2-IMMERSE

- The Architecture of the Service Development and Delivery Platform needs to be scalable to support the size of broadcast audiences (millions).
- Object-based models of content must be developed to support the level of flexible composition required to be responsive to needs across the broadcast value chain. If

experiences are to be built up from combinations of micro-services that can be recombined to support new content experiences the object model must not be genre specific or the addition of genre-specific plug-ins should be supported.

- Composition of content and micro-services needs to be flexible to work across a multiplicity of content micro-service combinations curated and selected by users across many and different devices and in some cases across locations. In at least two of our pilots we will connect locations to enhance the social experience in one case and learning in another.
- Synchronisation should be appropriate for each combination of micro-services (depending on the needs and latency tolerances of the audience. Synchronisation should also be consistent whether it's between micro-services within or across locations.
- By enabling all the stakeholders across the broadcast value chain to affect the selection and layout of micro-services the impact of changes in the balance of curation and customization must be understood. Designing the pilots with stakeholders and audience feedback will help us do that.

4. Methods

To help with the design of the Service Development and Delivery Platform and the Service Pilots to be trialed we will use a number of methods. These activities will involve designers, social scientists developers and stakeholders and include:

- Design workshops

  Design workshops usually take place early in the process and will be reported in the service pilot specific deliverables in this deliverable series (e.g. the trajectories workshop for Theatre in the Home).

- Wireframes of the layouts and interaction designs

  To facilitate the discussion of the control, content layout and interaction surfaces between stakeholders, designers, experimenters and developers we will create a series of 'wireframes' representing up the control and content surfaces layout and the interaction flows.

- Prototyping testing

  Prototype components of the service pilots will be tested in the lab by the developers. We include in this document an example of early bench prototyping to test the available audio solutions for our video chat, micro-service that will be included in our theatre service pilots.

- User Studies

  Where user studies: surveys, user lab testing or observational ethnography are carried out these will be reported in this series of deliverables. In addition to in the moment interactions by audiences during trials we will study and consider the needs for new craft skills within the production workflow required to deliver the new immersive environments we envision. WP3 and WP5 will combine to deliver this strand. Ethnographic observation interviews and design testing will inform the design of production tools.

Below, we describe our experiments on sychronisation and tolerance for latency for theatre and sport micro-services and audio testing for video-chat. In the theatre scenario we intend to phase the video chat in and out so as not to disturb engagement with the performance. This will allow us to develop and test the platform's timeline component and test the value of theatre ritual for remote viewing of a performance. Both sychronisation and video-chat have relevance to more than one service pilot.

5. Companion Screen Content: Tolerance for Latency

BBC R&D initiated a research theme to ascertain the boundaries of synchronisation delays, users perceive and tolerate in a companion screen experience. The questions within the theme considered what the minimal perceivable delay, between content being presented on a companion screen device (tablet) and content being presented on a TV, might be. Even if users could perceive the delay, at what point does it start getting annoying or distracting. To invert the question, how tolerant are our audience (users) to different levels of synchronisation accuracy?

Previously, BBC R&D, with partners in the DVB Project [14], developed open standards for the network protocols needed for a TV and a companion device to communicate – the Companion Screens and Streams specification (DVB-CSS) published as an ETSI Technical Specification [15]. The specification includes protocols that enable a TV to tell a companion device 'what content it is presenting' and 'what the timeline position is' thereby supporting reliable, accurate, and timely interactions over the home network. This allowed BBC R&D to build a companion screen experience prototype which allowed frame accurate synchronisation between the tablet and the TV [16].

Given the numerous factors which might have an impact on user experience in a companion screen enhanced programme, the study focused on content classed under the "Factual", "Sports" and "Drama" genres. Stimuli was prepared to represent three hypothetical companion screen experiences based on the most likely to be useful form factor given the genre (type of programme) being tested. The three different experiences all involved a user being asked to watch a clip from a previously broadcasted video, in addition to viewing suitable/relevant content on a tablet. Efforts were made to ensure that the companion content of production quality.

The companion screen experiences used were:

- A sports video on video use case in which users, watching clips of 'tries' from the "England versus France Six Nations 2015 final" rugby game [17] on the TV, were presented a supposedly synchronised video from an alternative camera angle of the same 'try' on the tablet. The alternative camera angle clips were sourced from the BBC sports library. The duration of the clips used ranged between 30 seconds and 1 minute based on how fast the players scored a try.
- A drama audio description on video use case in which users, watching clips from the 5th episode of the six-part period show "Wolf Hall" [18] on the TV, could also listen to the associated audio description stream on the tablet. The audio description clip was stripped from an archive of the broadcasted transport stream. The duration of the clips were about 2 minutes long and presented a complete scene within the drama.
- A factual web-based slide show on video use case in which users, watching clips from the 'winter' episode of the three-part natural history documentary "Alaska: Earth's Frozen Kingdom" [19] on the TV. The images and text were sourced by a freelance programme editor and the slideshow was designed by a UX designer from the BBC iPlayer team. The clips were 2.5 minutes long and presented a complete scene within the programme. New slides (an image with some text) were presented at about 30 second intervals. In a perfectly synchronised condition, a change in slide occurred during a scene cut on the TV.

After taking into consideration, current guidelines on A/V sync and preliminary results from pilots using a wider range of synchronisation delay values, the study focused on seven levels of synchronisation delay. The control condition had no delays injected into the companion screen experience prototype. The seven delays were injected deliberately into the system to simulate a delayed tablet. Three types of companion screen experiences and eight levels of delays gave rise to 24 distinct conditions.

- Video on video use case with synchronisation delays at 0, 0.02, 0.05, 0.1, 0.2, 0.5, 1 and 2 seconds
- Audio description on video use case with synchronisation delays at 0.1, 0.2, 0.5, 1, 2, 3 & 4 seconds
- Slide show on video use case with synchronisation delays at 0.1, 0.2, 0.5, 1, 2, 3 & 4 seconds

Thirty-two (16 male and 16 female) participants were recruited using a London-based agency with provisions to ensure that the participants were not biased against watching sports, drama and factual (particularly natural history documentary) content. Participants were also screened to ensure they were not biased against the BBC (or its programme offerings). Using a modified Latin square design, participants were asked to 'watch' twelve clips of varying types of experience and levels of synchronisation delays in a pre-set random order. At the end of the study, each distinct condition (clip of specific experience and level of synchronisation delay) was viewed sixteen times. Each participant took about ninety minutes to finish one session of viewing and gauging twelve clips. This included a ten-minute break in the middle of the session.

On the day, just before starting their study, each participant gave informed consent, answered a demographics questionnaire, and a questionnaire to gauge their media habits. Participants were given a sample run through of the three different types of experiences (in control conditions). They were then told they would be gauging twelve companion screen experiences. After viewing each experience, they were asked to fill a questionnaire gauging the experience. The questionnaire asked participants to agree or disagree on thirty statements along a 7-point Likert type scale. After watching all twelve conditions assigned to them, the participant was asked to complete a post-study questionnaire gauging companion screen experiences in general. A short semi-structured interview was then used to get qualitative responses from the participant. Finally, participants were debriefed. Participants were videotaped (including audio) through the study session.

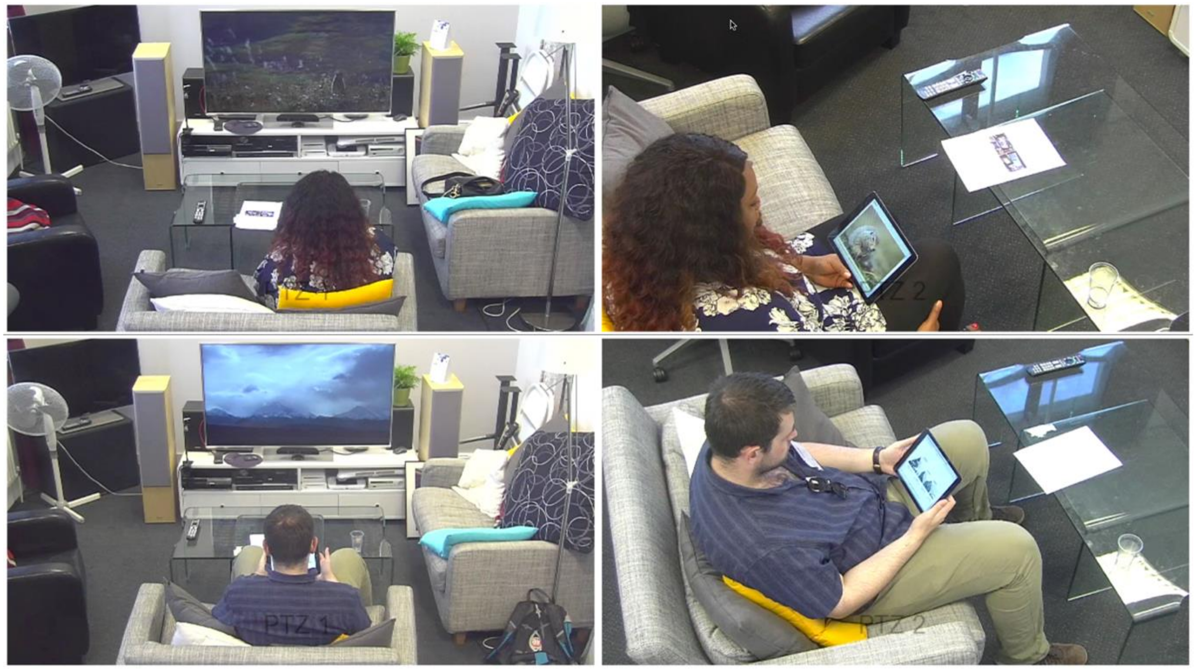


Figure 3. Apparatus set-up

Preliminary analysis of some of the data collected indicated that participants felt that the different types of experiences failed at being ‘well-timed’ at different levels of synchronisation delays. Participants were unable to distinguish synchronisation delays less than 0.5 seconds while watching a video on video companion screen experience presenting short sequences of a relatively fast paced sports programme. This value increased to 1 second when participants listened to audio description on the tablet while watching a slow paced period drama on the TV. In the case of looking at a slide show on the tablet while watching a natural history documentary on the TV, participants failed to notice synchronisation delays up to 4 seconds.

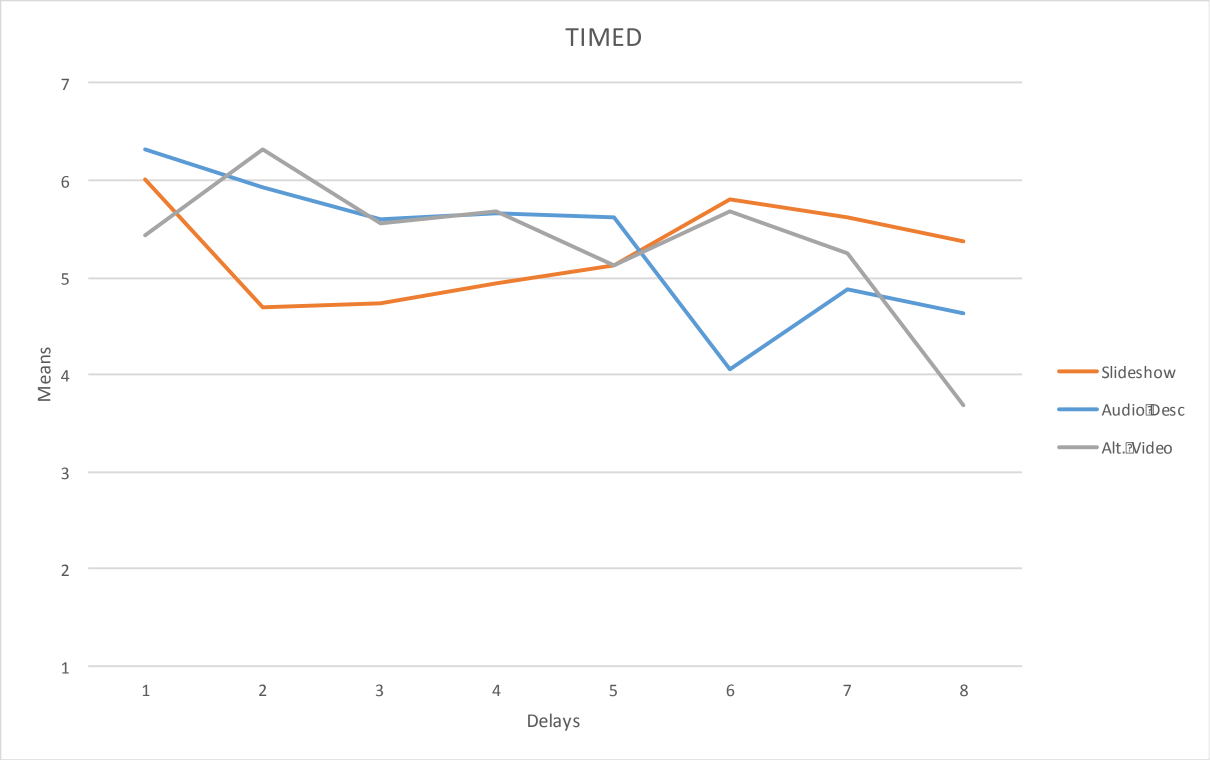


Figure 4: Mean response scores (1-7): I felt the TV and the Tablet were well timed to each other.

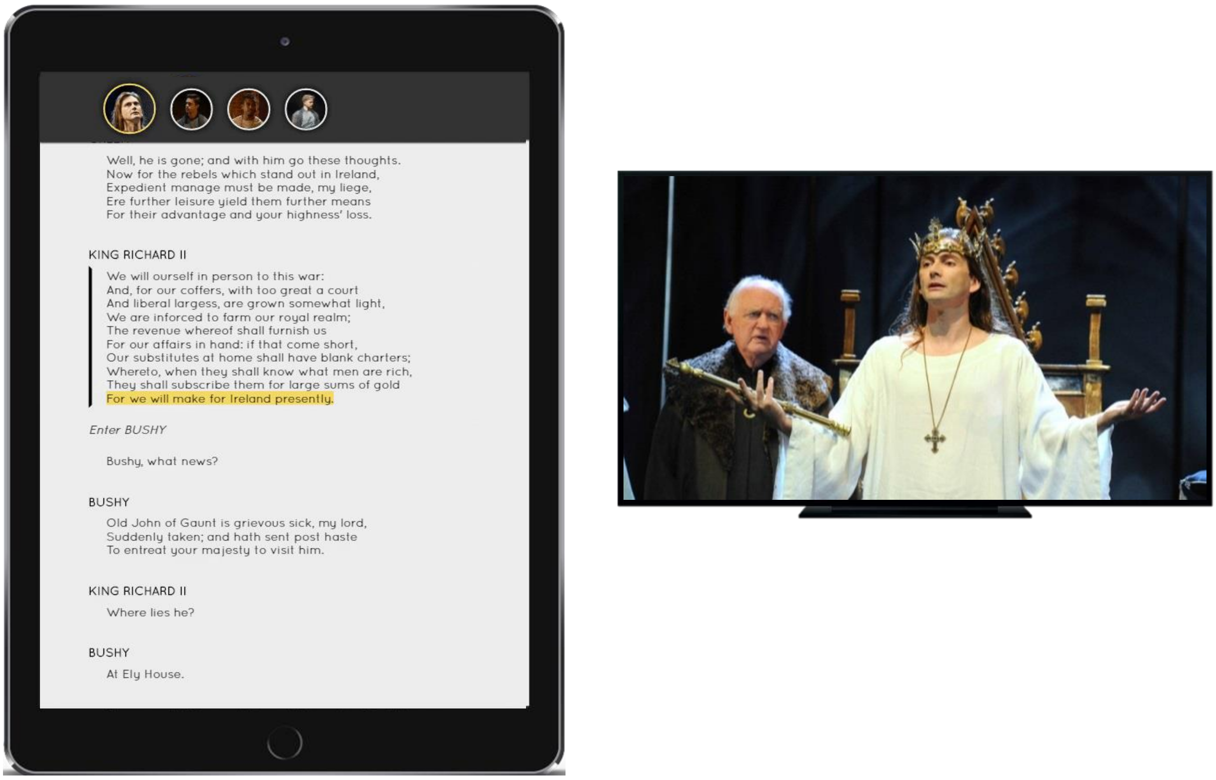
Detailed analysis of the data is ongoing and will be published at TVX 2017.

‘Shakespeare Synchronized Transcript Prototype’ by BBC & IRT

In addition to the more generalised design of the study, BBC R&D and IRT collaborated to conduct a more focused trial using one of the content forms closely coupled to the 2-IMMERSE use case. In this case, there were also three different types of experiences. However, the differentiation between the types of experiences were made using the type of interaction (passive, exploratory & call-to-action) afforded to the participants.

Like the generalised study, participants were also subjected to different levels of synchronisation delays. Again the choice of delay levels was selected after pilots were run with colleagues within BBC R&D. In the pilot, we tested delays in the order of magnitude of perceived lip sync (±0.05 and ±0.1 seconds) [14], of the average speech rate (±0.2 and ±0.5 seconds) [20] as well as values in the order of magnitude of accepted delays for TV subtitles (±1 and 2 seconds) [21,22]. Positive delays indicate that the tablet content lags behind the TV content while negative values indicate that tablet content is ahead. As only a few of the participants noticed the delays in lip-synchronism range, the delay values ±0.05 and ±0.1 seconds were excluded from the set of factor levels to study in the main experiment. In order to reduce the size of the experiment, the largest negative and the largest positive delays were removed. The levels of delays tested were -0.5, -0.25, 0, 0.25, 0.5, 1 seconds.

A prototype was built to present a synchronised transcript of Richard II, a Shakespearean play [23] alongside a video of the play on the TV. As depicted in Figure , the prototype presented a transcript of the play. The prototype had three modes of operation – passive, exploratory & call-to-action (Figure 6).



Well, he is gone; and with him go these thoughts.
Now for the rebels which stand out in Ireland,
Expedient manage must be made, my liege,
Ere further leisure yield them further means
For their advantage and your highness' loss.

KING RICHARD II

We will ourself in person to this war:
And, for our coffers, with too great a court
And liberal largess, are grown somewhat light,
We are inforced to farm our royal realm;
The revenue whereof shall furnish us
For our affairs in hand: if that come short,
Our substitutes at home shall have blank charters;
Whereto, when they shall know what men are rich,
They shall subscribe them for large sums of gold
For we will make for Ireland presently.

Enter BUSHY

Bushy, what news?

BUSHY

Old John of Gaunt is grievous sick, my lord,
Suddenly taken; and hath sent post haste
To entreat your majesty to visit him.

KING RICHARD II

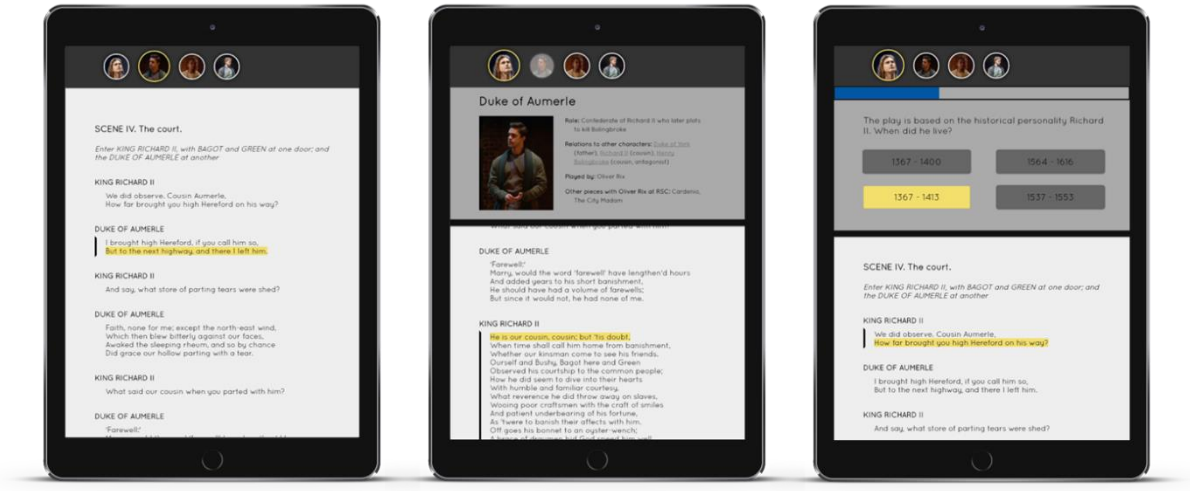
Where lies he?

BUSHY

At Ely House.

Figure 5. Media synchronisation between Tablet & TV

In the passive mode, the prototype merely presents the synchronised transcript as described above with no means of user interaction. The prototype runs at the speed dictated by the TV allowing the user to take in both pieces of content as the experience progresses. Under a perfectly synchronised condition, the current line being enacted in the play would be highlighted in yellow in addition to the current paragraph being scored using a vertical black runner line on the left hand side of the ‘page’. In addition, the corresponding character icon displayed across the top of the companion screen prototype bar would be highlighted with a thin yellow border to indicate which character was speaking the line.



SCENE IV. The court.

Enter KING RICHARD II, with BAGOT and GREEN at one door; and the DUKE OF AUMERLE at another

KING RICHARD II

We did observe. Cousin Aumerle,
How far brought you high Hereford on his way?

DUKE OF AUMERLE

I brought high Hereford, if you call him so,
But to the next highway, and there I left him.

KING RICHARD II

And say, what store of parting tears were shed?

DUKE OF AUMERLE

Faith, none for me; except the north-east wind,
Which then blew bitterly against our faces,
Awaked the sleeping rheum, and so by chance
Did grace our hollow parting with a tear.

KING RICHARD II

What said our cousin when you parted with him?

DUKE OF AUMERLE

'Farewell:'

Duke of Aumerle

Role: Confederate of Richard II who later plots to kill Bolingbroke

Relations to other characters: Duke of York (father), Richard II (cousin), Henry Bolingbroke (cousin, antagonist)

Played by: Oliver Rix

Other pieces with Oliver Rix at RSC: Cardenio, The City Madam

DUKE OF AUMERLE

'Farewell:'
Marry, would the word 'farewell' have lengthen'd hours
And added years to his short banishment,
He should have had a volume of farewells;
But since it would not, he had none of me.

KING RICHARD II

He is our cousin, cousin; but 'tis doubt,
When time shall call him home from banishment,
Whether our kinsman come to see his friends.
Ourself and Bushy, Bagot here and Green
Observed his courtship to the common people;
How he did seem to dive into their hearts
With humble and familiar courtesy,
What reverence he did throw away on slaves,
Wooing poor craftsmen with the craft of smiles
And patient underbearing of his fortune,
As 'twere to banish their affects with him.
Off goes his bonnet to an oyster-wench;

The play is based on the historical personality Richard II. When did he live?

SCENE IV. The court.

Enter KING RICHARD II, with BAGOT and GREEN at one door; and the DUKE OF AUMERLE at another

KING RICHARD II

We did observe. Cousin Aumerle,
How far brought you high Hereford on his way?

DUKE OF AUMERLE

I brought high Hereford, if you call him so,
But to the next highway, and there I left him.

KING RICHARD II

And say, what store of parting tears were shed?

Figure 6. Passive, Exploratory & Call-to-Action Modes

In the exploratory mode, the character icons on the top part of the companion screen prototype are clickable. On clicking, the user is presented with an information drawer in which information about the character is displayed. The information displayed included the name of the character, a photograph of the character, the role of the character in the current play, the relationships maintained between the character and others in the same play, the name of the actor portraying the character and other Royal Shakespearean Company productions the actor has starred in. Users can click on the other characters in the information drawer and explore character relationships during the experience.

In the call-to-action mode, a flash notification appears on the bottom right corner of the TV to indicate activity on the companion screen prototype. In a perfectly synchronised situation, the user is presented with a quiz question related to Richard II and Shakespeare in general. Once the question is presented, users have twenty seconds (as indicated through a progress bar) to choose an answer out of four possible responses.

Similar to the generalised study discussed in the previous sub-section, the study was a repeated-measures within-participant design. Unlike the generalised study, all participants in this study were exposed to all eighteen (18) conditions. Each participant was presented three blocks of six clips grouped by type of interaction. Within each block of six, the participant experience clips of the same interaction type with all six levels of delay counterbalanced using a 6x6 Latin square design. Similarly, the three blocks of six were varied using a 3x3 Latin square design.

Eighteen (9 male and 9 female) participants were recruited. They all had an affinity for Shakespearean plays and/or the theatre in addition to not being biased against the BBC. Participants went through an experimental procedure similar to the one described in the generalised study. Each participant took about a hundred and twenty minutes to finish one session of viewing and gauging eighteen conditions. This included a fifteen-minute break after experiencing twelve of the eighteen conditions.

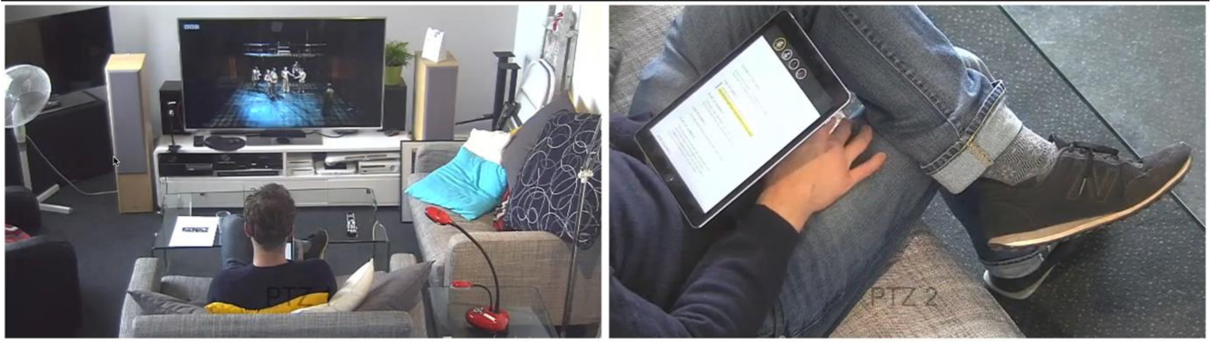


Figure 7. Apparatus set-up

Eighteen participants each experienced eighteen conditions and filled in a 16-item questionnaire after each condition resulting in 5184 responses. 'Interaction logs' were recorded to capture how participants interacted with the companion screen prototype as well as video footage of the experiences. Gaze behaviour was extracted from the video footage to be used as a form of objective behavioural response. Additionally, each participant was interviewed after the study.

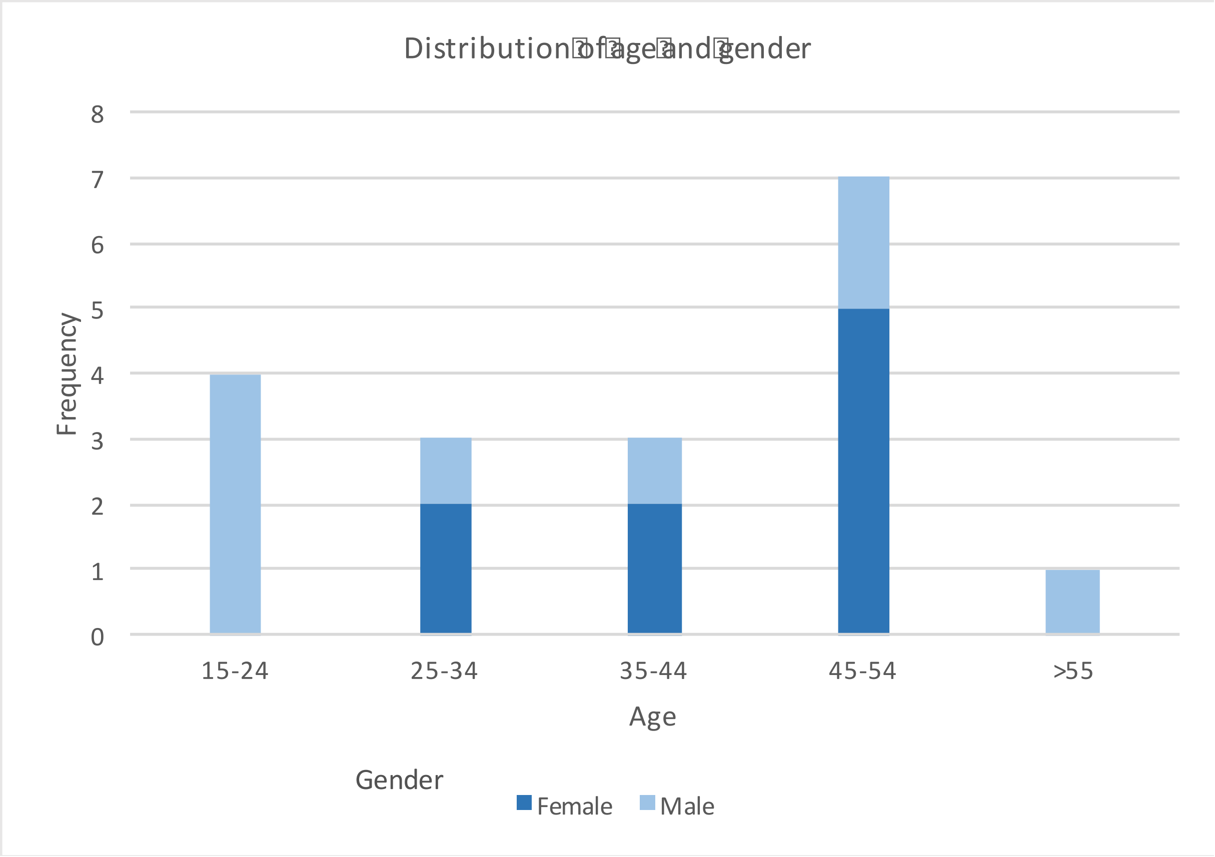


Figure 8 Demographics of Participants

Participants were between 18 and 55 years of age and gender balanced. All participants stated that they used a secondary device, such as a smart phone, tablet, and/or laptop, to search for information related to the TV programme. They were familiar with the technology used in the experiment. A majority self-assessed themselves as frequent users of touch-screen devices and computers.

In addition to the gaze behaviour data, a subset of the questionnaire responses was analysed. Questions cover perception of delays Q1 to Q3 and attention split Q4 and Q5 between companion and TV:

- Q1 - I felt the TV and the Tablet were well timed to each other.
- Q2 - I felt like I was waiting for the TV to "catch up" to the Tablet.
- Q3 - I felt like I was waiting for the Tablet to "catch up" to the TV.
- Q4 - I felt I was missing part of the programme on the TV because of the Tablet.
- Q5 - I felt I was missing content on the Tablet because of the TV.

Single-factor analysis was conducted across the responses for the different conditions. If results of Shapiro-Wilk tests indicated that a normal distribution of the samples could not be assumed, a non-parametric Friedman test was used to detect influences of participant responses on the conditions experienced. If the Shapiro-Wilk test showed that the samples comes from a normal distributed population, then a repeated-measures single-factor analysis of variance (rANOVA) was applied to analyse the influence on the response.

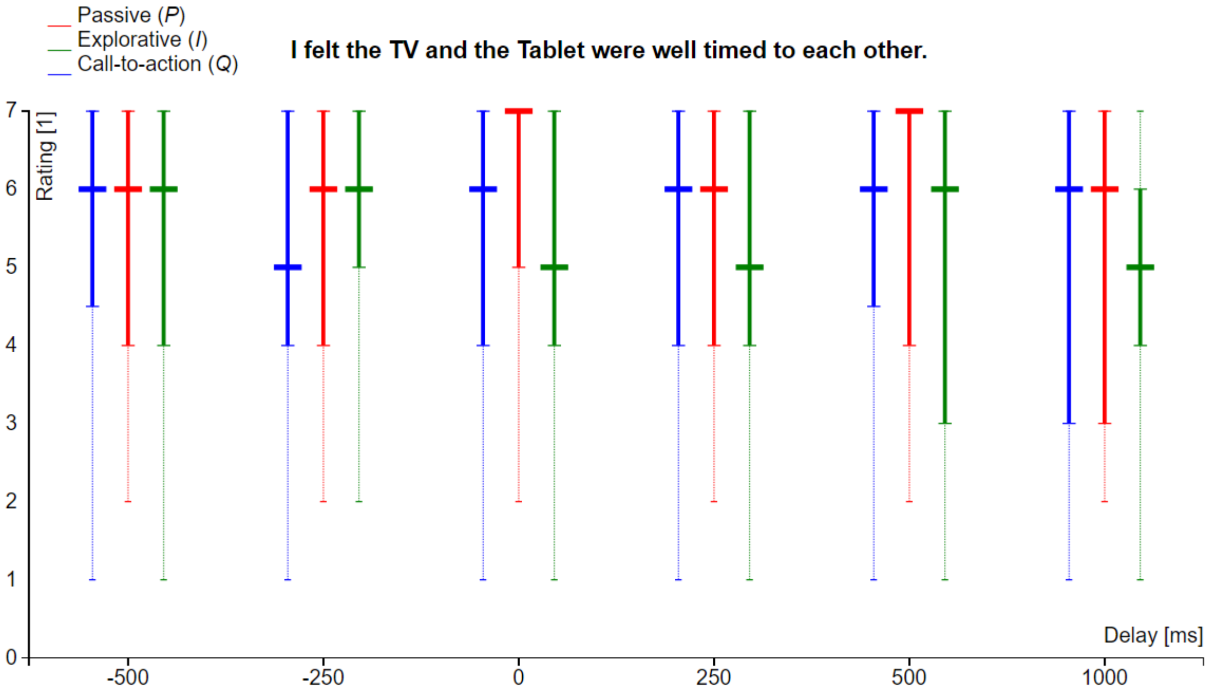


Figure 9. Responses for Q1

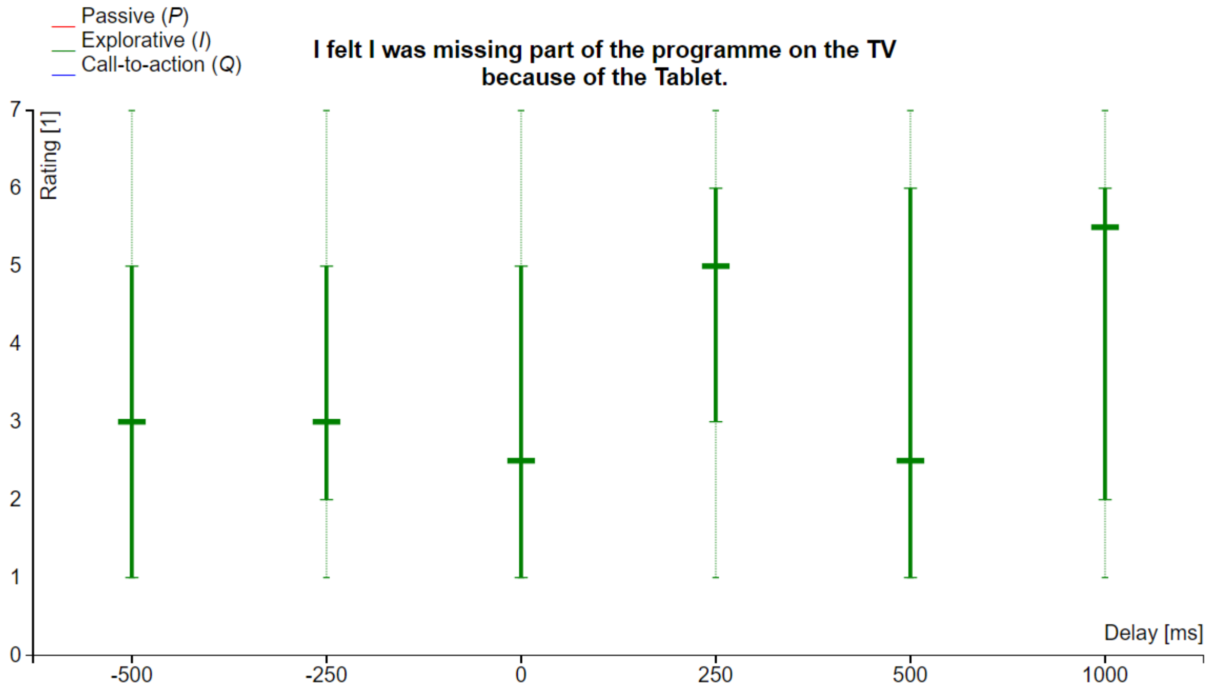


Figure 10. Responses for Q4

No significant effects were detected in participant responses to question Q1 to Q3 across the different delays for any of the chosen interaction levels. As can be observed in
Figure , the means are all quite high so it is very likely that, for the type of companion screen experience tested in this study, users will not notice synchronisation delays between -0.5 and 1 second. Significant effect was detected (Figure), across the time delay, when users were able to actively browse content on the companion screen (interaction level - exploration). During the conditions in which participants were able to actively browse additional content on the companion-screen, users' felt more distracted by the tablet at a delay of 1 second as opposed to lower delay values. This indicated that delays can affect aspects of the user experience before users start noticing them.

These results require that errors introduced during production of content timings and errors introduced by the synchronisation system do not accumulate to a value that exceeds the investigated delay range. Manual crafting of timings for the Shakespearean play transcript, used in the study, was comparatively time consuming. A production process that is at least partially automated will be a more practicable and efficient process. Future work could investigate efficient ways to generate the timing information, in particular, the applicability of professional subtitling tools or natural language processing tools. Future designs of studies should include evaluative methods that focus on more than the users' perception of delays.

Analysis and findings from part of the data gathered from this study will be published in TVX 2017. The data collected during the experiment has potential for further investigation. For instance, the interview recordings can be used to assess the value (utility & enjoyment) of the companion screen prototype through thematic analysis. Any findings will be published and data will be contributed to the Open Data Pilot.

6. Audio Performance: The Challenge of Integrating Video-Chat

To explore the practicalities and technical demands of video chat in the home pilots we have considered and tested solutions for use in an audio environment that comprises connected multi-screen homes. Audio isolation through beam-forming and multi-source echo cancelation tests are informing both the audio subsystem and the choice of HBBTV 2.0 set-top box and smart TV emulator.

Our goal was to enable open microphone multi-participant voice chat between living rooms whilst watching theatre on television.

The technical problems to over come are:

1. Reverberation in the room
    a. Varied room configurations
    b. Changes when people move about
2. Sound coming from the loudspeakers
    a. Picked-up by the microphone
    b. Mix of voice chat and programme output
    c. Multi-channel (stereo, 5.1 etc.)
    d. Synchronised programme output in multiple homes
3. Dynamic range of the sounds to handle
4. Noise: comparable to voice level at 3.5 meters

The Xbox Kinect provides a solution for open microphone chat in multi-player games that also allows speech recognition to work effectively when voices are captured in noisy environments. Microsoft researchers developed several key technologies to help them achieve this:

1. Microphone enclosure design (1.1dB improvement to directivity index)
2. Multichannel Acoustic Echo Cancelation (reduces 15-20dB echo)
3. Adaptive beamformer (reduces 3-6dB noise)
4. Spatial filtering / sound source localisation (reduces 6-12dB noise)

The challenge is to harness as many of these techniques as possible for the theatre home trial, which has much in common with the multi-player gaming scenario.

6.1 Phones and Tablets

If tablets are used as a lighter weight alternative (to a Kinect-like device) for video chat, users would find it hard to get a decent (and non-distracting) image to be transmitted to the remote parties unless we forced people to stand the tablets upright on a coffee table (which they may not have, or may not be in the right position). A natural behaviour is to put the tablet on a lap or arm of chair while watching television resulting in an image of the ceiling.

If people are using their own devices we would expect a range of different device capabilities, so we can't guarantee multi-microphone echo cancellation. We conducted a small test involving a Skype call between an iPad 2 and an Android tablet at separate locations. Each device was positioned ~2.5m from a TV playing the same broadcast simultaneously at a comfortable level. In spite of the adaptive echo control Skype implements, both tablets picked up bursts of TV audio and transmitted it with delay and distortion which was enough to be annoying and make conversation difficult.

It would be difficult to implement a system that automatically reduces the level of the theatre audio based on people’s chat interactions without having first removed/reduced background audio. Without this, the risk of false triggers would be very high and this could badly impact the effectiveness of the programme.

For tablet’s to work as voice chat clients, we would propose the following specific features:

1. A simple UI button on the tablet that allows one home to say something to everyone else. Pressing the button would unmute their tablet’s microphone and reduce the theatre audio levels on their TV. It could also potentially automatically trigger the subtitling DMApp component to appear (either on TV or device) so that they didn’t miss any key lines.
2. Allow microphone muting and theatre audio level reductions to be authored in the timeline document. By doing this we could identify in advance key events during the play when mutual applause or other audience reactions are expected and the social rituals allowed to flow more naturally.

The camera positioning issue above could be mitigated by having a central fixed webcam (attached to the TV emulator) which is only used for video capture, with audio capture and playback for chat being carried by a different device – so in effect we could have two completely separate WebRTC video and audio sessions if we wanted to, although this adds complexity.

6.2 Voice Activity Detection

In a conversation, people usually take it in turns to speak, however Hamlet will talk over you as his voice comes out of your television. This is distracting and makes it hard to hold a conversation.

One solution is to dip the volume of the programme when people speak using push-to-talk button(s) or voice activity detection (VAD). VAD would be more natural in terms of social interactions, but push-to-talk may be more practical. Dipping of volume would need to be triggered by both local and remote push-to-talk buttons (or by both local and remote microphones if VAD is used).

False positives, such as detecting Hamlet’s voice should are mitigated by performing VAD after cancellation of unwanted programme audio from the microphone input. The WebRTC code base features a VAD implementation that can also be used standalone.

6.3 Acoustic Echo Cancellation

Sound captured by the microphone contains several undesirable signals.

- Background noise
- Programme audio from the TV loud speaker
- Reverberation due to reflected sound

  - Generated by people talking and the output from the loud speaker
- Echoes and feedback due to hearing yourself on the TV loud speaker

Noise varies over time and is subject to environmental changes. Adaptive noise suppression is required to track and eliminate noise. WebRTC provides noise suppression features.

Hamlet’s voice being played out of a television speaker is equivalent to another person in the room. We would want his voice to be removed by acoustic echo cancellation (AEC) and not delivered to the far-end living room.

AEC is an adaptive algorithm requiring both microphone input and speaker output. It needs low-level access to the speaker output(s) to successfully cancel other sounds generated by the television. Further analysis is required to determine whether WebRTC AEC will cancel all audio generated by the speaker, or just audio from the remote microphone. AEC is also highly sensitive to clock drift, which may play an important role in the success of the cancellation.

6.4 Beam forming

Beam forming dramatically cuts out sound emitted by the television speakers. In our tests, this was seen with the Samson beam forming mic, the Playstation 3 Eye and Logitech c910 web cameras.

These devices use a broadside microphone array in which the line of microphones is arranged perpendicular to the preferred direction of sound waves.

A disadvantage of broadside arrays is that they only attenuate sound coming from the side of the array; the attenuation is the same in front and behind the array. To reject sound from the rear Microsoft opted to use 4 individual supercardioid microphones in a broadside configuration. Cardioid/supercardioid microphones also do a good job of rejecting sound from the rear.[1]

We considered using Kinect for Windows as a target platform, but unfortunately all audio DSP processing in Kinect for Windows, including beam forming, has been removed in the latest release (v2).

The Kinect Sensor will give us a beam direction and confidence for the loudest sound, but doesn't give us access to the processed output.

6.5 WebRTC/getUserMedia

Google Chrome introduced beam forming for WebRTC recently and it is configured using the getUserMedia() constraints object, where the microphone array geometry must be specified manually.

Google have been working on this since mid 2015. It means that Chrome now has an equivalent audio processing capability to Microsoft's Voice Capture DSP. A possible next step would be to enable these experimental features in our WebRTC test app to turn the beam forming on/off. The caveat is that multi-channel acoustic echo cancellation is not supported yet in WebRTC and beam forming is only enabled by default on ChromeOS.

6.6 PulseAudio

Beam forming is now available as a module for Pulse Audio on Linux. See:

https://arunraghavan.net/2016/06/beamforming-in-pulseaudio/

This module wraps Chrome's WebRTC audio processing capability and has been adapted to support multi-channel acoustic echo cancellation. In our tests, beam forming worked well, but further work is required to configure and evaluate echo cancellation performance. Using this module in conjunction with WebRTC means disabling WebRTC's audio processing features in Chrome. An advantage is that PulseAudio is more readily configurable and works at the low-level.

PulseAudio must be configured to remap the two centre channels of the microphone array to arrange the inputs linearly for the beam former. The beam forming worked exceptionally well in our tests using Audacity, as did the stereo microphones of the Logitech c910.

6.7 Lab Testing

Lab testing was performed to understand the technical constraints on the audio element of the video chat solution and to look at the acceptability of the performance achieved.

We looked at three different video chat configurations:

1. Simple 2-way video chat (one room at Adastral Park and one at CWI in Amsterdam)

2. 3-way video chat (adding a second isolated room at Adastral Park)

3. The user experience of moving between pre-recorded content and video chat.

In each room we had a TV Emulator device (either a desktop PC or Odroid device). A large TV with built-in speakers and a good quality webcam were connected. We also used a keyboard and mouse for control, although in a trial we would expect either an infra-red remote control or interaction via a companion device. We used the Google Chrome browser on the desktop PC and Chromium on the Android-powered Odroid.

We concluded that the video chat experience can be good enough for the theatre at home scenario, on the basis that it is enabled before and after the play and during the interval, and that no pre-recorded content (video or audio) is played out while video chat is active. We observed the audio quality for video chat when using several different webcams and also with a separate USB microphone, and we tried changing the distance between these and the chat participants. As could be expected, the best audio quality is achieved when the microphone is closest to those speaking (ie. <1m away). Conversely, the best place for the camera is either above or below the TV screen.

Multi-party audio chat with open microphones and speakers is challenging, and browsers such as Chrome and Firefox support a number of different WebRTC audio processing features which are intended to improve the experience, such as echo cancellation, noise suppression and automatic gain control. While we have limited control over these (basically 'on' or 'off'), we found that enabling all of them gave the best experience.

Adding more endpoints (going from 2-way to 3-way chat and beyond) is more challenging for echo cancellation but we found 3-way chat to be acceptable as long as the microphones were close to the participants.

We explored the user experience of being able to switch between a screen containing pre-recorded material (in our case a synopsis of the play with an audio recording from Radio 4 playing in the background) and 3-way video chat. The UI enabled the remote participants to be seen (but not heard) as thumbnails next to the synopsis text. Each endpoint has the option to enable their microphone if they want to say something to the others, and if they want to have a conversation they can click to switch between this screen and full video chat. This flexibility seems to work well, as long as the button functions are made clear. We also considered that the control buttons could be made available on a companion device instead of the main TV screen.

6.8 Audio Chat Conclusion

There have been significant developments in bringing Kinect-like audio processing capabilities to WebRTC and PulseAudio. More work is required to fully evaluate these audio stacks, but it's clear that they don't deliver the quality we require for 2-IMMERSE unless certain compromises are made. All the audio processing technologies are available, but haven't been pulled together in one place in a way that's readily exploitable. For example, WebRTC beam forming is still disabled on most major platforms and it doesn't appear to support beam steering, sound source localization, multi-channel echo cancellation or provide a means of calibrating echo cancellation.

There are very few low-cost USB microphone array devices out there. Expensive solutions are available for professional conferencing applications, but then there is a big gap, followed by games console devices like Kinect. Laptops quite often have 2x microphone arrays, but they are only suitable when users are sat a couple of feet away from the screen.

Multiplayer computer gaming has generated the demand for devices like Kinect, but until 2-IMMERSE, there hasn't really been a demand for equivalent open-microphone chat in a synchronised multi-home broadcast television experience. Consequently, there is a shortage of cheap USB microphone array devices that integrate with WebRTC and are configured to work over a large range of distances. This presents a possible future market opportunity.

In summary, for the theatre at home trial we will:

- Capture video using a cheap webcam attached to the TV Emulator.
- Render chat video either on the TV or companion device.
- Render chat audio on the TV (possibly in mono to assist echo cancellation in Chrome)
- Use the WebRTC echo cancellation functions to allow open-microphone chat in the same way as Skype or FaceTime.

- Use the flexibility of our DMApp environment and Timeline service to control the companion device microphone and theatre audio levels on the TV to minimise the chances of the latter being picked up by the companion device.

7 Summary

The 2-IMMERSE project proposes to create an extensible platform to deliver new media services that deliver flexibly configured multi-screen experiences. Unlike previous projects delivering multi-screen content, the 2-IMMERSE platform will support the orchestration of media across devices by stakeholders across the broadcast value chain. This will allow producers, broadcasters, venue owners and audiences to select content micro-services we call Distributed Media Applications (DMApps) and their distribution across devices. The challenge can be summarised as co-producing the preferred content for the preferred screen at the right time whilst allowing a satisfying mix of control across the value chain.

Experiments, design studies and technology research are contributing to our understanding of the challenges and potential solutions. By engaging producers, broadcasters, venue owners and potential audiences in the conception and testing of our micro-services Distributed Media Apps) we will create and test a scalable and extensible platform for developing tailored and immersive engagement with content and data. The relationships between people, content, devices and locations set the challenges and drive the extended capabilities of the 2-IMMERSE platform. Each service pilot has the potential to exercise these capabilities and test their value.

The results of synchronisation and latency tolerance will be published at TVX 2017. The search for an audio chat solution beyond the current default solution will continue until the trial code freeze for the first pilot. Our investigation into new production craft and workflows will start in earnest in the run-up to the first sports trial. Technical use cases will continue to emerge out of the technology platform development. These will be captured in subsequent documents. Latest documentation: https://2-IMMERSE.eu/wiki/technical-use-cases/

A Note on Production Craft

The first pilot: Theatre in the Home has been driven by 2-IMMERSE partner Illuminations who produce the broadcasts of Royal Shakespeare company performances. The emphasis for this pilot will be on the audience experience leaving the production workflow unchanged but for the provision of a full stage view for selection by the audience at home at any time. Distributed Media Apps build features around the broadcast of the performance such as a scrolling script, access to video chat, views of the theatre lobby and online information about the performers and the play.

Future pilots will learn more about the production workflows and craft in the provision of sports coverage which is richer in camera viewpoints, simultaneous action and the generation of data.

On a broader note the aim of the project is to build an extensible platform that enables more than the four pilots covered during the project lifetime. Our intention is that other genres and live events such as music, and the athletics could be covered by adding DMApps to the library available to cover other facets of these experiences not present in our exemplars.

8 References

[1] Pablo Cesar, Dick CA Bulterman, and AJ Jansen. Usages of the secondary screen in an interactive television environment: Control, enrich, share, and transfer television content. In Changing television environments, pages 168–177. Springer, 2008.

[2] Santosh Basapur, Gunnar Harboe, Hiren Mandalia, Ashley Novak, Van Vuong, and Crysta Metcalf. Field trial of a dual device user experience for itv. In Proceddings of the 9th interna- tional interactive conference on Interactive television, pages 127–136. ACM, 2011.

[3] Santosh Basapur, Hiren Mandalia, Shirley Chaysinh, Young Lee, Narayanan Venkitaraman, and Crysta Metcalf. Fanfeeds: evaluation of socially generated information feed on second screen as a tv show companion. In Proceedings of the 10th European conference on Interactive tv and video, pages 87–96. ACM, 2012.

[4] Mark Lochrie and Paul Coulton. Sharing the viewing experience through second screens. In Proceedings of 10th European Conference on Interactive TV & Video, pp 199–202. ACM, 2012. [5] X-Factor App www.itv.com/xfactor/app

[6] BBC Internet Blog. New antiques roadshow play-along app. www.bbc.co.uk/ariel/20692943 2015-03-04.

[7] Channel 4 Press. Million pound drop mobile app hits over 1 million. http://www.channel4.com/info/press/news/million-pound-drop-mobile-app-hits-over-1-million. Accessed: 2014-10-05.

[8] Janet Murray, Sergio Goldenberg, Kartik Agarwal, Tarun Chakravorty, Jonathan Cutrell, Abraham Doris-Down, and Harish Kothandaraman. Story-map: ipad companion for long form tv narratives. In Proceedings of the 10th European conference on Interactive tv and video, pages 223–226. ACM, 2012.

[9] Michael E Holmes, Sheree Josephson, and Ryan E Carney. Visual attention to television

programs with a second-screen application. In Proceedings of the Symposium on Eye Tracking Research and Applications, pages 397–400. ACM, 2012.

[10] Vinoba Vinayagamoorthy, Penelope Allen, Matt Hammond, and Michael Evans. Researching the user experience for connected tv: a case study. In CHI'12 Extended Abstracts on Human Factors in Computing Systems, pages 589–604. ACM, 2012.
[11] Andy Brown, Michael Evans, Caroline Jay, Maxine Glancy, Rhianne Jones, and Simon Harper. HCI over multiple screens. In CHI'14 Extended Abstracts on Human Factors in Computing Systems, pages 665–674. ACM, 2014.

[12] Dagmar Kern, Paul Marshall, and Albrecht Schmidt. Gazemarks: gaze-based visual placeholders to ease attention switching. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 2093–2102. ACM, 2010.
[Survey results] Guardian Apps blog: Social TV and second-screen viewing: the stats in 2012 www.theguardian.com/technology/appsblog/2012/oct/29/social-tv-second-screen-research

[13] MediaScape EU FP7 (610404) Dynamic Media Service Creation http://mediascapeproject.eu/about.html

[14] The DVB Project – about page. DVB. Retrieved September 24, 2015 from https://www.dvb.org/about.

[15] Digital Video Broadcasting (DVB); Companion Screens and Streams; Part 1: Concepts, roles and overall architecture. 2015. DVB BlueBook A167-1, ETSI TS 103 286-1. Retrieved September 29, 2015 from https://www.dvb.org/standards.

[16] Vinoba Vinayagamoorthy, Rajiv Ramdhany, and Matt Hammond. 2016. Enabling Frame-Accurate Synchronised Companion Screen Experiences.
In Proceedings of the ACM International Conferenceon Interactive Experiences for TV and OnlineVideo (TVX '16). ACM, New York, NY, USA, 83-92.
http://dx.doi.org/10.1145/2932206.2932214.

[17] Six Nations 2015: England 55 – 35 France – BBC Sport. BBC Sports pages. Last accessed 15th March 2017.
http://www.bbc.co.uk/sport/rugby-union/31974720

[18] BBC Two - Wolf Hall. BBC /programme pages. Last accessed 15th March 2017.
http://www.bbc.co.uk/programmes/p02gfy02

[19] BBC Two – Alaska: Earth's Frozen Kingdom. BBC /programme pages. Last accessed 15th March 2017. http://www.bbc.co.uk/programmes/b0520nyz

[20] J. Yuan, M. Liberman, and C. Cieri. 2006. Towards an integrated understanding of speaking rate in conversation. In *Interspeech 2006*. 541–544.

[21] D. Burnham, J. Robert-Ribes, , and R. Ellison. 1998. Why captions have to be on time. In Audio-visual speech processing. 153–156.

[22] Ichiro Maruyama, Yoshiharu Abe, Eiji Sawamura, Tetsuo Mitsuhashi, Terumasa Ehara, and Katsuhiko Shirai. 1999. Cognitive experiments on timing lag for superimposing closed captions. In Sixth European Conference on Speech Communication and Technology. 575–578.

[23] BBC – Shakespeare lives, Richard II. BBC /programme pages. Last accessed 16th March 2017. http://www.bbc.co.uk/programmes/p03rr1v1