

UNIVERSIDADE DE LISBOA

Faculdade de Ciências

Departamento de Informática



SENSING AND AWARENESS OF 360°
IMMERSIVE VIDEOS ON THE MOVE

João Carlos Reis Ramalho

DISSERTAÇÃO

MESTRADO EM ENGENHARIA INFORMÁTICA

Engenharia de Software

2013

UNIVERSIDADE DE LISBOA

Faculdade de Ciências

Departamento de Informática



SENSING AND AWARENESS OF 360°
IMMERSIVE VIDEOS ON THE MOVE

João Carlos Reis Ramalho

DISSERTAÇÃO

MESTRADO EM ENGENHARIA INFORMÁTICA

Engenharia de Software

Trabalho orientado pela Prof. Doutora Maria Teresa Caeiro Chambel

2013

Acknowledgements

First and foremost, I want to thank my Professor and advisor, Prof. Teresa Chambel, for guiding me throughout this journey. The creative freedom that she entrusted in me really opened the window for me to get the required inspiration to produce something that I can say I am proud of and can call my own. Also, I am grateful for the endless patience, immense knowledge, valuable advices, critical opinion, and for the countless brainstorm sessions that I enjoy so much.

I would like to thank Fundação para a Ciência e Tecnologia for the financial support through the research project (UTA- Est/ MAI/0010/2009) (url-ImTV). I extend this appreciation to LaSIGE and the Department of Informatics of FCUL for the outstanding conditions provided to accomplish the work comprised in this thesis. In the scope of the ImTV project, an especial thanks to the ImTV project colleagues of Universidade Nova de Lisboa for the collaboration regarding the Facial Expression Recognition Framework.

To those friends always caring. You know who you are.

To Joana, for growing up beside me. May I thank you for being unconditionally supportive and for putting up with me every single day (which is a mighty task in itself). Thank you for sharing countless moments of joy with me, but, most importantly, for always being able to turn moments of sorrow in moments of pure happiness.

I cannot express in words, nor in any other way, how thankful I am to my parents. I can only thank you for the permanent guidance, continuous love and unconditional support, moral values and education. I am today, and will continue to be, what you made of me. Also, I must never forget to express my deep gratitude towards you for introducing me to the greatest passion of my life: music. Pedro, I thank you for always being able to put up with everything that means being the younger brother ;).

Finally, to my grandfather, for being my role model, and my best friend throughout the forever-short period of life I got to share with you.

To my grandfather, for showing me the true meaning of life.

Resumo

Ao apelar a vários sentidos e transmitir um conjunto muito rico de informação, o vídeo tem o potencial para causar um forte impacto emocional nos espectadores, assim como para a criação de uma forte sensação de presença e ligação com o vídeo. Estas potencialidades podem ser estendidas através de percepção multimídia, e da flexibilidade da mobilidade. Com a popularidade dos dispositivos móveis e a crescente variedade de sensores e actuadores que estes incluem, existe cada vez mais potencial para a captura e visualização de vídeo em 360° enriquecido com informação extra (metadados), criando assim as condições para proporcionar experiências de visualização de vídeo mais imersivas ao utilizador.

Este trabalho explora o potencial imersivo do vídeo em 360°. O problema é abordado num contexto de ambientes móveis, assim como num contexto da interação com ecrãs de maiores dimensões, tirando partido de *second screens* para interagir com o vídeo. De realçar que, em ambos os casos, o vídeo a ser reproduzido é aumentado com vários tipos de informação. Foram assim concebidas várias funcionalidades para a captura, pesquisa, visualização e navegação de vídeo em 360°. Os resultados confirmaram a existência de vantagens no uso de abordagens multisensoriais como forma de melhorar as características imersivas de um ambiente de vídeo. Foram também identificadas determinadas propriedades e parâmetros que obtêm melhores resultados em determinadas situações.

O vídeo permite capturar e apresentar eventos e cenários com grande autenticidade, realismo e impacto emocional. Para além disso, tem-se vindo a tornar cada vez mais pervasivo no quotidiano, sendo os dispositivos pessoais de captura e reprodução, a Internet, as redes sociais, ou a iTV exemplos de meios através dos quais o vídeo chega até aos utilizadores (Neng & Chambel, 2010; Noronha et al, 2012). Desta forma, a imersão em vídeo tem o potencial para causar um forte impacto emocional nos espectadores, assim como para a criação de uma forte sensação de presença e ligação com o vídeo (Douglas & Hargadon, 2000; Visch et al, 2010). Contudo, no vídeo tradicional a experiência dos espectadores é limitada ao ângulo para o qual a câmara apontava durante a captura do vídeo. A introdução de vídeo em 360° veio ultrapassar essa restrição.

Na busca de melhorar ainda mais as capacidades imersivas do vídeo podem ser considerados tópicos como a percepção multimídia e a mobilidade. Os dispositivos móveis têm vindo a tornar-se cada vez mais omnipresentes na sociedade moderna, e, dada a grande variedade de sensores e actuadores que incluem, oferecem um largo espectro de oportunidades de captura e reprodução de vídeo em 360° enriquecido com informação extra (metadados), tendo portanto o potencial para melhorar o paradigma de interação e providenciar suporte a experiências de visualização de vídeo mais ponderosas e imersivas. Contudo, existem desafios relacionados com o design de ambientes eficazes que tirem partido deste potencial de imersão.

Ecrãs panorâmicos e CAVEs são exemplos de ambientes que caminham na direção da imersão total e providenciam condições privilegiadas no que toca à reprodução de vídeo imersivo. Porém, não são muito convenientes e, especialmente no caso das CAVEs, não são facilmente acessíveis. Por outro lado, a flexibilidade associada aos dispositivos móveis poderia permitir que os utilizadores tirassem partido dos mesmos usam-do-os, por exemplo, como uma janela (móvel) para o vídeo no qual estariam imersos. Mais do que isso, seguindo esta abordagem os utilizadores poderiam levar estas experiências de visualização consigo para qualquer lugar. Como *second screens*, os dispositivos móveis podem ser usados como auxiliares de navegação relativamente aos conteúdos apresentados no ecrã principal (seja este um ecrã panorâmico ou uma CAVE), representando também uma oportunidade para fazer chegar informação adicional ao utilizador, eliminando do ecrã principal informação alheia ao conteúdo base, o que proporciona uma melhor sensação de imersão e flexibilidade.

Este trabalho explora o potencial imersivo do vídeo em 360° em ambientes móveis aumentado com vários tipos de informação. Nesse sentido, e estendendo um trabalho anterior (Neng, 2010; Noronha, 2012; Álvares, 2012) que incidiu maioritariamente na dimensão participativa de imersão, a presente abordagem centrou-se na dimensão perceptual de imersão. Neste âmbito, foram concebidas, desenvolvidas e testadas várias funcionalidades, agrupadas numa aplicação de visualização de vídeo em 360° – Windy Sight Surfers. Considerando a crescente popularidade dos dispositivos móveis na sociedade e as características que os tornam numa oportunidade para melhorar a interação homem-máquina e, mais especificamente, suportar experiências de visualização de vídeo mais imersivas, a aplicação Windy Sight Surfers está fortemente relacionada com ambientes móveis. Considerando as possibilidades de interação que o uso de *second screens* introduz, foi concebida uma componente do Windy Sight Surfers relacionada com a interação com ecrãs de maiores dimensões.

Os vídeos utilizados no Windy Sight Surfers são vídeos em 360°, aumentados com uma série de informações registadas a partir do Windy Sight Surfers durante a sua captura. Isto é, enquanto a câmara captura os vídeos, a aplicação regista informação adicional – metadados – obtida a partir de vários sensores do dispositivo, que complementa e enriquece os vídeos. Nomeadamente, são capturadas as coordenadas geográficas e a velocidade de deslocamento a partir do GPS, a orientação do utilizador a partir da bússola digital, os valores relativos às forças-G associadas ao dispositivo através do acelerómetro, e são recolhidas as condições atmosféricas relativas ao estado do tempo através de um serviço web. Quando capturados, os vídeos, assim como os seus metadados, podem ser submetidos para o sistema.

Uma vez capturados e submetidos, os vídeos podem ser pesquisados através do mais tradicional conjunto de palavras chave, de filtros relacionados com a natureza da aplicação (ex. velocidade, período do dia, condições atmosféricas), ou através de um mapa, o que introduz uma componente geográfica ao processo de pesquisa. Os resultados podem ser apresentados numa convencional lista, no formato de uma cover-flow, ou através do mapa.

No que respeita à visualização dos vídeos, estes são mapeados em torno de um cilindro, que permite representar a vista dos 360° e transmitir a sensação de estar parcialmente rodeado pelo vídeo. Uma vez que a visualização de vídeos decorre em dispositivos móveis, os utilizadores podem deslocar continuamente o ângulo de visão do vídeo 360° para a esquerda ou direita ao mover o dispositivo em seu redor, como se o dispositivo se tratasse de uma janela para o vídeo em 360°. Adicionalmente, os utilizadores podem alterar o ângulo de visualização arrastando o dedo pelo vídeo, uma vez que todo o ecrã consiste numa interface deslizante durante a visualização de vídeos em 360°.

Foram ainda incorporadas na aplicação várias funcionalidades que pretendem dar um maior realismo à visualização de vídeos. Nomeadamente, foi desenvolvido um acessório de vento na plataforma Arduino que leva em conta os metadados de cada vídeo para produzir vento e assim dar uma sensação mais realista do vento e da velocidade do deslocamento durante a visualização dos vídeos. De referir que o algoritmo implementado leva em conta não só a velocidade de deslocamento, como também o estado do tempo em termos de vento (força e orientação) aquando da captura do vídeo, e a orientação do utilizador de acordo com o ângulo do vídeo a ser visualizado durante a reprodução do vídeo.

Considerando a componente áudio dos vídeos, neste sistema, o áudio de cada vídeo é mapeado num espaço sonoro tridimensional, que pode ser reproduzido num par de auscultadores estéreo. Neste espaço sonoro, a posição das fontes sonoras está associada

ao ângulo frontal do vídeo e, como tal, muda de acordo com o ângulo do vídeo a ser visualizado. Isto é, se o utilizador estiver a visualizar o ângulo frontal do vídeo, as fontes sonoras estarão localizadas diante da cabeça do utilizador; se o utilizador estiver a visualizar o ângulo traseiro do vídeo, as fontes sonoras estarão localizadas por de trás da cabeça do utilizador. Uma vez que os vídeos têm 360°, a posição das fontes sonoras varia em torno de uma circunferência à volta da cabeça do utilizador, sendo o intuito o de dar uma orientação adicional no vídeo que está a ser visualizado.

Para aumentar a sensação de movimento através do áudio, foi explorado o Efeito de Doppler. Este efeito pode ser descrito como a alteração na frequência observada de uma onda, ocorrendo quando a fonte ou o observador se encontram em movimento entre si. Devido ao facto deste efeito ser associado à noção de movimento, foi conduzida uma experiência com o intuito de analisar se o uso controlado do Efeito de Doppler tem o potencial de aumentar a sensação de movimento durante a visualização dos vídeos. Para isso, foi adicionada uma segunda camada sonora cuja função é reproduzir o Efeito de Doppler ciclicamente e de forma controlada. Esta reprodução foi relacionada com a velocidade de deslocamento do vídeo de acordo seguinte proporção: quanto maior a velocidade, maior será a frequência com que este efeito é reproduzido.

Estas funcionalidades são relativas à procura de melhorar as capacidades imersivas do sistema através da estimulação sensorial dos utilizadores. Adicionalmente, o Windy Sight Surfers inclui um conjunto de funcionalidades cujo objectivo se centra em melhorar as capacidades imersivas do sistema ao providenciar ao utilizador informações que consciencializem o utilizador do contexto do vídeo, permitindo assim que este se aperceba melhor do que se está a passar no vídeo. Mais especificamente, estas funcionalidades estão dispostas numa camada por cima do vídeo e disponibilizam informações como a velocidade atual, a orientação do ângulo do vídeo a ser observado, ou a força-G instantânea. A acrescentar que as diferentes funcionalidades se dividem numa categoria relativa a informação que é disponibilizada permanentemente durante a reprodução de vídeos, e numa segunda categoria (complementar da primeira) relativa a informação que é disponibilizada momentaneamente, sendo portanto relativa a determinadas porções do vídeo.

Procurando conceber uma experiência mais envolvente para o utilizador, foi incorporado um reconhecedor emocional baseado em reconhecimento de expressões faciais no Windy Sight Surfers. Desta forma, as expressões faciais dos utilizadores são analisadas durante a reprodução de vídeos, sendo os resultados desta análise usados em diferentes funcionalidades da aplicação. Presentemente, a informação emocional tem três aplicações no ambiente desenvolvido, sendo usada em: funcionalidades de

catalogação e pesquisa de vídeos; funcionalidades que influenciam o controlo de fluxo da aplicação; e na avaliação do próprio sistema.

Considerando o contexto do projeto de investigação ImTV (url-ImTV), e com o intuito de tornar a aplicação o mais flexível possível, o Windy Sight Surfers tem uma componente *second screen*, permitindo a interação com ecrãs mais amplos, como por exemplo televisões. Desta forma, é possível utilizar os dois dispositivos em conjunto por forma a retirar o melhor proveito de cada um com o objectivo de aumentar as capacidades imersivas do sistema. Neste contexto, os vídeos passam a ser reproduzidos no ecrã conectado, ao passo que a aplicação móvel assume as funcionalidades de controlar o conteúdo apresentado no ecrã conectado e disponibilizar um conjunto de informações adicionais, tais como um minimapa, onde apresenta uma projeção planar dos 360° do vídeo, e um mapa da zona geográfica associada ao vídeo onde se representa o percurso em visualização em tempo real e percursos adicionais que sejam respeitantes a vídeos associados à mesma zona geográfica do vídeo a ser visualizado no momento.

Foi efectuada uma avaliação de usabilidade com utilizadores, tendo como base o questionário USE e o Self-Assessment Manikin (SAM) acoplado de dois parâmetros adicionais relativos a presença e realismo. Com base na observação durante a realização de tarefas por parte dos utilizadores, foram realizadas entrevistas onde se procurou obter comentários, sugestões ou preocupações sobre as funcionalidades testadas. Adicionalmente, a ferramenta de avaliação emocional desenvolvida foi utilizada de forma a registar quais as emoções mais prevalentes durante a utilização da aplicação. Por fim, as potencialidades imersivas globais do Windy Sight Surfers foram avaliadas através da aplicação do Immersive Tendencies Questionnaire (ITQ) e de uma versão adaptada do Presence Questionnaire (PQ). Os resultados confirmaram a existência de vantagens no uso de abordagens multisensoriais como forma de melhorar as características imersivas de um ambiente de vídeo. Para além disso, foram identificadas determinadas propriedades e parâmetros que obtêm melhores resultados e são mais satisfatórios em determinadas condições, podendo assim estes resultados servir como diretrizes para futuros ambientes relacionados com vídeo imersivo.

Palavras-chave: Vídeo, Imersão, Presença, Percepção Visual, 360°, Percepção Tátil, Vento, Percepção Auditiva, Áudio 3D, Mobilidade, Movimento, Velocidade, Orientação, Emoção, Segundo Ecrã.

Abstract

By appealing to several senses and conveying very rich information, video has the potential for a strong emotional impact on viewers, greatly influencing their sense of presence and engagement. This potential may be extended even further with multimedia sensing and the flexibility of mobility. Mobile devices are commonly used and increasingly incorporating a wide range of sensors and actuators with the potential to capture and display 360° video and metadata, thus supporting more powerful and immersive video user experiences.

This work was carried out in the context of the ImTV research project ([url-ImTV](#)), and explores the immersion potential of 360° video. The matter is approached in a mobile environment context, and in a context of interaction with wider screens, using second screens in order to interact with video. It must be emphasized that, in both situations, the videos are augmented with several types of information. Therefore, several functionalities were designed regarding the capture, search, visualization and navigation of 360° video. Results confirmed advantages in using a multisensory approach as a means to increase immersion in a video environment. Furthermore, specific properties and parameters that worked better in different conditions have been identified, thus enabling these results to serve as guidelines for future environments related to immersive video.

Keywords: Video, Immersion, Presence, Perception, Visual Sensing, 360°, Tactile Sensing, Wind, Auditory Sensing, 3D Audio, Mobility, Movement, Speed, Orientation, Emotion, Second Screen.

Contents

Chapter 1	Introduction	1
1.1	Motivation	2
1.2	Research Objectives	5
1.3	Context	6
1.4	Contributions	7
1.5	Development Plan	8
1.6	Document Structure	10
Chapter 2	State of the Art.....	13
2.1	Immersion.....	13
2.2	Historical Context of Immersion in Video	17
2.3	Hypervideo	23
2.4	360° Hypervideo	27
2.5	Wind Based Interfaces.....	30
2.6	3D Audio	33
2.7	Emotions.....	36
2.7.1	Emotional Models	36
2.7.2	Emotional Representations and Visualizations	38
2.8	Social TV.....	41
2.9	Second Screen	46
2.9.1	TV Environments	47
2.9.2	Video Game Environments	49
2.9.3	Other Environments.....	51
2.10	Maps and Georeferenced Guidance.....	53
2.11	Recommendation Systems.....	56
Chapter 3	Windy Sight Surfers	59
3.1	Requirements Specification.....	60
3.1.1	Functional and Non-Functional Requirements.....	60
3.1.2	Use Cases.....	62

3.2	User Registration and Authentication.....	65
3.3	Video & Metadata Capture.....	65
3.4	Video & Metadata Sharing.....	70
3.5	Video Search.....	71
3.6	Perceptual Sensing Features	76
3.6.1	Visual Sensing in 360° Video.....	76
3.6.2	Tactile Sensing Through a Wind Accessory	77
3.6.3	Auditory Sensing: Spatial Audio.....	80
3.6.4	Auditory Sensing: Cyclic Doppler Effect.....	81
3.7	Context Awareness Features	83
3.7.1	Permanent Information Features:	83
3.7.2	Momentary Information Features:.....	84
3.8	The Emotional Perspective.....	86
3.8.1	Video Emotional Cataloguing	87
3.8.1.1	EmoMap.....	87
3.8.1.2	EmoMe.....	89
3.8.2	Emotion Driven Control Flow.....	89
3.8.3	User Evaluation Based on Emotional Impact.....	90
3.9	Interaction with TVs & Wider Screens using Second Screens	90
3.9.1	Minimap	91
3.9.2	Hyperlinks	91
3.9.3	Geographical Navigation & Orientation in the 360° Videos and Maps	92
3.9.4	Related Videos.....	92
Chapter 4	System Implementation	93
4.1	System Architecture	93
4.2	System Development Methodology	97
4.3	Metadata Capture.....	98
4.3.1	GPS Tracking Component.....	99

4.3.2	Orientation Tracking Component.....	99
4.3.3	Weather Status Tracking Component.....	100
4.3.4	G-Force Tracking Component.....	102
4.4	Location Based Video Search.....	103
4.5	Perceptual Sensing Features	105
4.5.1	Visual Sensing in 360° Video	105
4.5.2	Wind Accessory.....	106
4.5.3	Auditory Sensing	111
4.6	Emotional Perspective	114
4.7	Interaction with TV's & Wider Screens	115
4.7.1	Controlling the TV's Video Viewing Angle Through the Mobile Application	116
4.7.2	Geographical Navigation & Orientation in the 360° Videos and Maps	116
Chapter 5	User Evaluation	119
5.1	Method.....	120
5.2	Results	122
5.2.1	Video Search Features	122
5.2.2	Perceptual Sensing Features	123
5.2.3	Context Awareness Features	127
5.2.4	Perceptual Sensing & Context Awareness features in Conjunction.	129
5.2.5	Emotional Features	130
5.2.6	Interaction with TV's & Wider Screens Features	131
5.2.7	Evaluating the Emotional Impact of Windy Sight Surfers	133
5.2.8	Global Presence and Immersion Evaluation.....	134
5.2.9	Final Overview	135
Chapter 6	Conclusions and Future Work	137
6.1	Conclusions	137
6.2	Future Work.....	139

Bibliography	141
Internet References	147
Annex A: Video's Metadata File Example	153
Annex B: User Evaluation One Script	155
Annex C: User Evaluation Two Script	157
Annex D: Immersive Tendencies Questionnaire	161
Annex E: Presence Questionnaire	165

List of Figures

Figure 1.1: Initial Gantt Chart	9
Figure 1.2: Final Gantt Chart.....	10
Figure 2.1: Representative image in the 5760x1080 format	15
Figure 2.2: Winky Dink and You	19
Figure 2.3: Content being streamed from an iPad to a TV through the AppleTV	21
Figure 2.4: Comparison between IMAX and conventional film systems	22
Figure 2.5: Hypersoap	24
Figure 2.6: HyperCafe	24
Figure 2.7: Hyper-Hitchcock.....	25
Figure 2.8: YouTube Annotations	26
Figure 2.9: 360° Hypervideo Player	29
Figure 2.10: Synchronization between 360° Hypervideo Player and maps	29
Figure 2.11: The WindCube.	31
Figure 2.12: Head Mounted Display	32
Figure 2.13: Lehmann notched box plot for presence in different wind prototypes	33
Figure 2.14: Emotional Models	37
Figure 2.15: Mappiness	38
Figure 2.16: Wee Feel Fine's different views	39
Figure 2.17: The Million Pound Drop's PlayAlong Application	41
Figure 2.18: Miso	42
Figure 2.19: TVCheck	43
Figure 2.20: Facebook's BuzMuzik Application	44
Figure 2.21: SentiTVchat with the sentiment graph.....	45
Figure 2.22: Avatar Theater.....	46
Figure 2.23: PDA prototype	47
Figure 2.24: Wii U GamePad	50
Figure 2.25: Xbox SmartGlass	51
Figure 2.26: GoPro's Mobile App.....	52

Figure 2.27: Akai's SysthStation keyboard controller	52
Figure 2.28: Line6 StageScape M20d	53
Figure 2.29: Google Maps Web Application	54
Figure 2.30: Google Maps Mobile Application	54
Figure 2.31: Nike+ Application.....	55
Figure 2.32: ATC9K Application.....	56
Figure 3.1: Use Cases Diagram	62
Figure 3.2: Sony Bloggie Handy Cam1	66
Figure 3.3: 360° Video, as captured by the Sony Bloggie camera	66
Figure 3.4: 360° Video, after being converted to a rectangle	66
Figure 3.5: Windy Sight Surfers: home screen (left) and Capture Mode (right).....	67
Figure 3.6: Windy Sight Surfers Capture Mode.....	69
Figure 3.7: Sharing the video's metadata (upper left) and the video itself (upper right); Notification for incomplete submissions (bottom).....	71
Figure 3.8: Windy Sight Surfers search through keywords and filters	72
Figure 3.9: Windy Sight Surfers Search Through Map.....	72
Figure 3.10: Search example.	73
Figure 3.11: Video Search results in Cover-flow presentation	74
Figure 3.12: Video search results in the map, with info-box highlighting information about one of the videos.....	75
Figure 3.13: Videos "Being Watched Now" functionality.....	75
Figure 3.14: Pan around the 360° video in both left and right directions by moving the tablet around	77
Figure 3.15: Drag interface.....	77
Figure 3.16: 3D Audio: Red rectangle representing the video viewing viewport.....	81
Figure 3.17: 3D Audio: source location changing around the 360° video viewing.....	81
Figure 3.18: Doppler Effect: Audio changes cyclically as in grey paths	82
Figure 3.19: Video Context Awareness: Video being reproduced with the overlay, displaying the orientation, speed and G-Force values.....	83
Figure 3.20: Different compass modes.....	84

Figure 3.21: Hyperlinks in 360° video.....	85
Figure 3.22: EmoMap.....	88
Figure 3.23: Windy Sight Surfers search through keywords and filters	88
Figure 3.24: Windy Sight Surfers Second Screen application	91
Figure 4.1: Windy Sight Surfers conceptual architecture.....	94
Figure 4.2: Windy Sight Surfers concrete architecture	95
Figure 4.3: Accelerometer's Axis.....	102
Figure 4.4: Android Activity's lifecycle.....	107
Figure 4.5: Wind Accessory inside view.....	108
Figure 4.6: Wind Accessory inside view.....	108
Figure 4.7: Wind Accessory's architecture (breadboard perspective)	109
Figure 4.8: Wind Accessory's architecture (schematic perspective)	109
Figure 4.9: Pulse Width Modulation	110
Figure 4.10: Waveforms used in sound design.....	113
Figure 5.1: Self-Assessment Manikin 9-Point Scale	121
Figure 5.2: Recognized Emotions during user Evaluation	133

List of Tables

Table 3.1: Beaufort Scale	79
Table 5.1: USE evaluation of the Video Search features	123
Table 5.2: SAM and PR evaluation of the Video Search features	123
Table 5.3: USE evaluation – Visual and Tactile	124
Table 5.4: SAM and PR evaluation – Visual and Tactile.....	124
Table 5.5: USE evaluation – Spatial Audio.....	125
Table 5.6: SAM and PR evaluation – Spatial Audio.....	125
Table 5.7: USE evaluation – Cyclic Doppler Effect	127
Table 5.8: SAM and PR evaluation – Cyclic Doppler Effect.....	127
Table 5.9: USE evaluation of Context Awareness features.....	128
Table 5.10: SAM and PR evaluation of Context Awareness features.....	128
Table 5.11: USE evaluation of Perceptual Sensing & Context Awareness features in Conjunction	129
Table 5.12: SAM and PR evaluation of Perceptual Sensing & Context Awareness features in Conjunction.....	130
Table 5.13: USE evaluation of Emotional features	131
Table 5.14: SAM and PR evaluation of Emotional features	131
Table 5.15: USE evaluation of Interaction with TV's & Wider Screens features.....	132
Table 5.16: SAM and PR evaluation of Interaction with TV's & Wider Screens features	133
Table 5.17: Immersive Tendencies Questionnaire	134
Table 5.18: Presence Questionnaire	134

Chapter 1

Introduction

Video allows capturing and presenting events and scenarios with great authenticity, realism, and emotional impact, and it is becoming pervasive in our lives, in personal capturing and display devices, over the Internet, in social media, and through video on demand services on iTV (Neng & Chambel, 2010; Noronha et al, 2012). With its immersive capabilities, video has a strong impact on the viewers' emotions, their sense of presence and engagement (Douglas & Hargadon, 2000; Visch et al, 2010). However, in traditional video, the user experience is limited to the angle that the camera was pointing during the video's capturing. 360° video is a technology that overcomes this restriction, and that has the potential to create highly immersive video scenarios, by allowing users to feel the experience of being surrounded by the video.

This potential may be extended even further with multimedia sensing and the flexibility of mobility. Mobile devices are commonly used and increasingly incorporating a wide range of sensors and actuators with the potential to capture and display 360° video enriched with additional information (metadata), thus supporting more powerful and immersive video user experiences. Nevertheless, there are challenges related to the design of effective environments that may profit from this immersive potential.

Widescreens and CAVEs, for their shapes and dimensions, are examples of steps towards full immersion, and are privileged displays for immersive video view. However, they are not very convenient, and especially CAVEs are not widely available. On the other hand, the flexibility associated with mobile devices could allow users to actually turn around, as if they hold in their hands a window to the video where they are immersed in. Moreover, mobile devices create the opportunity for users to bring this experience with them everywhere. As second screens, mobile devices may also be used to help navigation in a video that is projected or displayed outside in a wider screen (from TV to CAVE), allowing for example to display and interact with additional

information and navigational aids, thus clearing the video display of additional extraneous info, which contributes for an increased sense of immersion and flexibility.

This thesis addresses immersion in interactive environments for future TV and was carried out in the context of the ImTV research project ([url-ImTV](#)). It explores the immersive potential of 360° video enriched with additional information in mobile environments. This first chapter starts with the motivation behind this work and the main research objectives. It describes the scientific context in which design and development took place, presents the main contributions, alongside with the accomplished project's development plan and an overview of the thesis' structure.

1.1 Motivation

People strive for the fulfilment of their entertainment and information needs. When considering the technologies that support those needs, two technologies that are right on top of population's choices are TV and, more recently, the Internet. Since it first appeared, video has been one of the main means of communication, as it is able to present a large quantity of information in a rich cultural context through a brief period of time, and its effectiveness is well proven by its solid success over time. Television was the technology responsible for bringing video to the masses, and it has been one of the most popular means to deliver video ever since it first appeared. However, during the last decade, the world has seen an astonishing revolution due to the Internet's boom. In the late nineties, the Internet primarily consisted of a mean to communicate information, through simple websites, email clients or even social networks. But from then on, the Internet has grown immensely, broadening the content categories it provides and, due to technological development, allowing to store and transmit video more effectively. In particular, video consumption has been growing consistently, and rapidly (regarding both video submission and access). YouTube announced in 2013 one billion active users every month ([url-YoutubeReport](#)). This translates to one out of every two people on the Internet using YouTube, and therefore consuming video content. It is important to note that Internet's traffic is increasing exponentially every month, as stated by a Cisco study ([url-CiscoGlobalIPTraffic](#)), where the conclusions point to the fact that Global IP traffic has fourfold between 2012 and 2017, and is predicted to increase threefold over the next five years, being that, every second, 1.2 million minutes of video content is predicted to cross the network in 2016. Therefore, a conclusion that may be drawn is that video is already one of the main media types accessed on the Internet and its importance is still increasing at a high pace.

As an opportunity to integrate and extend the different media, thus enhancing their capabilities and flexibility, Interactive TV (iTV) is becoming a reality. Video has been

able to benefit greatly with the combination of TV and Internet, by becoming increasingly dominant in both means. Regarding iTV, this concept is often used interchangeably to describe a variety of rather different kinds of interactivity. However, these can be separated in three distinct interactivity categories. The simplest one refers to the interactivity with the TV set. This category of interaction started with the use of the remote control to enable channel surfing behaviours, and evolved to what is now known as video-on-demand. This kind of interaction does not change any content or its inherent linearity, only how users control the viewing of that content. Other category of interaction, and the one that ultimately represents the real “interactive TV” is the Interactivity with TV program content, which is also the most challenging to produce. This paradigm is based on the idea that the program itself might change based on the viewer’s input. The third interaction category, Interactivity with TV-related content, is commonly the least understood, and presents itself as the most promising when it comes to changing the way we watch TV over the next decade. Examples of this interaction level include getting more information about what is appearing on the TV, whether it is about complementing the news, additional information on sport events broadcasts, or publicity and e-commerce on clothes that actors wear in movies (Dakss, 1998; url-HyperSoap).

Immersion is related to the subjective experience of being fully involved in an environment or virtual world, and this concept is often associated with the concept of presence, which relates to the viewer’s conscious feeling of being inside the virtual world. When discussing immersive environments, a multitude of approaches that may enhance their immersive capabilities can be considered. Namely, technology has been playing a very important role, especially in the virtual reality field, as the computer games industry is becoming evermore interested in improving the realism of their approaches.

Recently, a new interest has arisen about the creation of immersive environments around video. The main objective is to offer a more immersive experience to users, so that they ultimately feel as part of that same environment, being 360° video a technology that is beginning to be used as a means to enhance immersion. Systems composed of wide screens or CAVES - provide privileged display conditions for immersive video viewing, but they are not very handy as they rely heavily on specific (and often expensive) hardware. CAVES in particular are not widely available. On the other hand, mobile devices are becoming ubiquitous in everyday life and represent, by the sensors and actuators they are increasingly incorporating, a wide range of opportunities to experiment new approaches. Their flexibility opens new opportunities for more realistic designs. Moreover, this is the first time an entire new generation has grown up accessing content on demand, which is characterized by the Internet, Mobile

and Social environments. Nielsen ([url-GenC](#)) defines this group of people by the connected behaviour, referring to it as the Generation-C. In this context, video appears as one of the most pervasive content types, and its consumption is ever increasing. Considering YouTube's report ([url-YoutubeReport](#)), although its numbers highlight the popularity of video in this group, they must be contextualized, by specifying what types of devices is Generation-C using to consume video content. Addressing this question, Google ([url-GoogleGenC](#)) states that YouTube usage on smartphones mirrors usage on PCs, and around 70% of the population watch YouTube on two or more devices, which confirms the significance of mobile devices in the video access context. Also, Strover et al. (Strover & Moner, 2012) investigated the role of Internet-based content, alongside various user-owned technologies (such as mobile devices and laptop computers), in order to map the new dynamics of entertainment media, with results suggesting that audiences are evolving to expect to create and use content in various forms and in various places. Strover concludes that using a mobile device, such as a laptop computer is nearly as common as using a television for entertainment. Considering the importance and proliferation of mobile devices in our society, there is an interest in finding out how they can contribute towards the enhancement of immersive characteristics in video environments. A possible approach considers the possibility to enhance the 360° video capture with metadata collected through the mobile device (e.g. geo-location, speed, or local weather), thus having the potential to support more powerful and immersive video user experiences. An example would be to allow users to actually turn around when viewing video, as if they held in their hands a window to the video where they were immersed in, while watching and sensing it, as if they were there. By using mobile platforms, users could bring this experience with them everywhere, benefiting from greater flexibility when compared to more immersive environments, such as CAVES, which impose significant hardware restrictions to users.

Also, mobile devices are being associated with the previously mentioned "Interactivity with TV-related content" interaction category. This is related with the recently popular concept of using mobile devices as second screens in TV or computer environments. Again regarding the video context, mobile devices may be used to assist navigation in a video that is being displayed in a wider screen (from TV to CAVE), allowing e.g. to display and interact with additional information and navigational aids. In addition to providing more information to the user, this approach allows clearing the video display of additional information, in order to reduce information overload and maximise the immersiveness of the environment.

The conjunction of all these concepts and technologies forms the guidelines for the proposed work.

1.2 Research Objectives

This project has as its main objective to explore means to increase immersion in video through the design and development of innovative interfaces that involve users in an interactive and immersive video environment using different media and devices. More specifically, immersion is addressed in a perceptual perspective, where the focus relates to the development of technologies that may enhance the users' video viewing and navigation experience, by striving to increase their sense of presence, and give them the feeling of being inside the video. Additionally, immersion is addressed in a participatory perspective, where users can contribute to the environment.

In the approach towards the described goal, where users are able to capture 360° video alongside additional information, referred to as the video's metadata, which is captured through a mobile device to enhance the 360° video with information regarding a variety of dimensions, such as geo-location, speed, or weather conditions. After capturing videos, users may submit them to the system (mainly targeted at mobile platforms), where there are a variety of search processes users can use in order to find videos in which they are interested (such as a set of "search through map" features). Focusing the design on the empowerment of users in their immersive video experiences, this approach may provide several components that enhance the video visualization by exploring senses or perceptions, which might increase the experience's realism and sense of presence. Namely, wind may be used in order to provide a more realistic perception of speed and movement. Additionally, sound is considered as a means to assist users' sense of orientation in the 360° video, and as a means to increase the movement sensation while viewing videos by the controlled use of a simulation of the Doppler Effect. Also, this system must in some way provide users with context awareness information, both to provide them with additional information about the video content (such as hyperlinks or information regarding perception – speed or weather information) and to alert to important information located outside the viewing area (such as link awareness, used to alert users to hyperlinks outside the viewing area). Beyond these functionalities, this system incorporates an emotional component, whose role in the environment is to provide users with a recommendation system based on emotional recognition. Lastly, this approach may allow the mobile application to be extended to iTVs and wider screens, thus creating a more immersive environment, where both categories of devices are used together. In this context, the work comprised in this thesis strives to answer the following Research Questions (RQs):

RQ1: "Do the designed map search features enhance the search process?"

RQ2: "Would a full screen pan-around interface increase the sense of immersion 'inside' the 360° video?"

RQ3: “Does wind contribute to increasing realism of sensing speed and direction in video viewing?”

RQ4: “Does a 3D mapping of the video sound allow for easier identification of the video orientation while it is being reproduced?”

RQ5: “Can a controlled use of the Doppler Effect increase the movement sensation while viewing videos?”

RQ6: “Do the designed context awareness features contribute to a more immersive environment?”

RQ7: “Do users consider the emotional perspective relevant in the access and search of videos?”

RQ8: “Does the interaction with TVs & wider screens, with video in full screen and additional content and navigation control in a second screen, contribute to a more immersive environment?”

Windy Sight Surfers was designed, developed and evaluated in order to address these research questions. This work was carried out in extension to a previous approach – Sight Surfers – in the ImTV research project (Noronha, 2012; Álvares, 2012).

1.3 Context

This work has been developed in the context of the *ImTV – “On-Demand Immersive TV for Communities of Media Producers and Consumers”* research project (UTA- Est/ MAI/0010/2009) (url-ImTV), in the Human-Computer Interaction and Multimedia (HCIM) research group (url-HCIM), which is part of the LaSIGE research lab of the Informatics Department of the Faculty of Sciences of the University of Lisbon (url-LaSIGE), in collaboration with the following entities: FCT-UNL, INESC-Porto, UTAustin - Communication and Information Schools, FCCN, ZON, RTP, Duvideo and MOG.

The ImTV project has as its main goal the design and development of an immersive and interactive TV environment that engages spectators and increases the overall quality of the video watching experience. In order to do so, the project envisions exploiting the full potential of new trends in media production and consumption by devising an on-demand immersive TV framework that combines the TV industry, Internet distribution models and end-user’s needs and interests.

1.4 Contributions

The major contributions of this work relate with Windy Sight Surfers: a prototype for a system that encompasses the capture, submission and visualization of 360° video and hypervideo, with several components aimed at the enhancement of the system's immersive capabilities in mind. In this context, the main solutions that were designed, implemented and evaluated include:

- The extension of a mobile application module for the capture and submission of video metadata (by extending the amount and categories of information acquired and standardizing the data that was already being captured) (Noronha, 2012).
- A set of search features to allow the users to search for videos according to keywords, filters, geographic locations, and emotional characteristics.
- A mobile application module for the reproduction of 360° videos, comprising different mechanisms to view around the different video angles (visual perceptual immersion), and a video overlay with a set of features designed to increase context awareness (using the video metadata as the source information).
- Extension of an algorithm for the synchronization between video and its metadata.
- A hardware prototype of a Wind Accessory for the enhancement of the speed and movement sensations while viewing 360° videos.
- A mobile application module for the reproduction of the video's audio component in a three dimensional space, as a means to enhance user orientation while viewing 360° videos.
- A mobile application module for the reproduction of a cyclic emulation of the Doppler Effect in a three dimensional space for the enhancement of the speed and movement sensations while viewing 360° videos.
- Extension of an auxiliary web interface for extending the mobile application to interactive TVs.
- A mobile interface aimed at the control of video reproduction through interactive TVs (exploring the second screen paradigm).

These contributions were presented to the scientific community through the publication of a full paper in the international reference conference on interactive TV and video ([url-EuroITV](#)), and a full paper in the Immersive Media Experiences workshop at ACM MM'13, the premier international conference on multimedia:

- Ramalho, J. and Chambel, T. "Immersive 360° Mobile Video with an Emotional Perspective". In Proceedings of the International ACM Workshop on Immersive Media Experiences 2013 at ACM Multimedia 2013, Barcelona, Spain, October 21-25th, 2013.
- Ramalho, J. and Chambel, T. "Windy Sight Surfers: sensing and awareness of 360° immersive videos on the move". In Proceedings of EuroITV 2013: 11th European conference on Interactive TV and video (pp. 107-116), Como, Italy, June 24-26th, 2013.

Still in the context of this work, I took part in additional activities during the year that allowed me to present my work and get more involved in the academic and research community and activities:

- Participation, as Web Chair, in the organization of the International ACM Workshop on Immersive Media Experiences 2013 ([url-ImmersiveME](#)) at the ACM Multimedia 2013 International Conference on Multimedia, Barcelona, Spain, October 21-25th, 2013 ([url-ACM](#)) (to be held), where I contributed with the design and development of the workshop website.
- Presentation of a Demo of the Windy Sight Surfers prototype in the “FCUL – Dia Aberto 2013” event, April 11th 2013 ([url-FCULDiaAberto](#)), where I presented my work and experience to younger students who are about to graduate from high school, and seeking information to choose a university graduation.
- Presentation of “Sharing & Navigating Immersive Video” with Prof. Teresa Chambel in the ImTV 2nd Workshop at INESC-Porto, January 10th 2013 ([url-ImTVWorkshop](#)).
- Participation in the 22^o Congresso das Comunicações da APDC 2012, November 21st-22nd 2013 ([url-APDC](#)) as a selected student in the “Talentos da Nova Geração” program.

1.5 Development Plan

Regarding the development plan, as the work comprised in this thesis has a strong focus on user interaction, it is imperative that its development methodology embraces the opportunities to introduce new elements that might arise during the course of the project. Therefore, the development was structured around an iterative methodology, where there was the possibility to add new functionalities and refine the ones being designed.

The first period of this work was focused on the exploration of concepts and related works on the areas more closely related to this thesis. This approach allowed to get a better understanding of the core themes of this work, and constituted the basis for the creative process of designing the features that would later result in this thesis.

Afterwards, focus shifted to the design and implementation of the referred features, which were comprised in the Windy Sight Surfers Application. These features were evaluated afterwards in order to analyse their perceived usability and usefulness.

Given the dynamic nature of this work, the iterative methodology that was adopted for this project, and the opportunities to publish my work earlier, the initial development plan suffered modifications, as can be seen in figures 1.1 and 1.2. Also, the project's planned duration of nine months was exceeded by two months, being the main motivations to this extension, the writing of two research papers, instead of one, as it was initially planned, the involvement on the Immersive Media Experiences ACM workshop organization, and the other reported activities that were not initially planned.

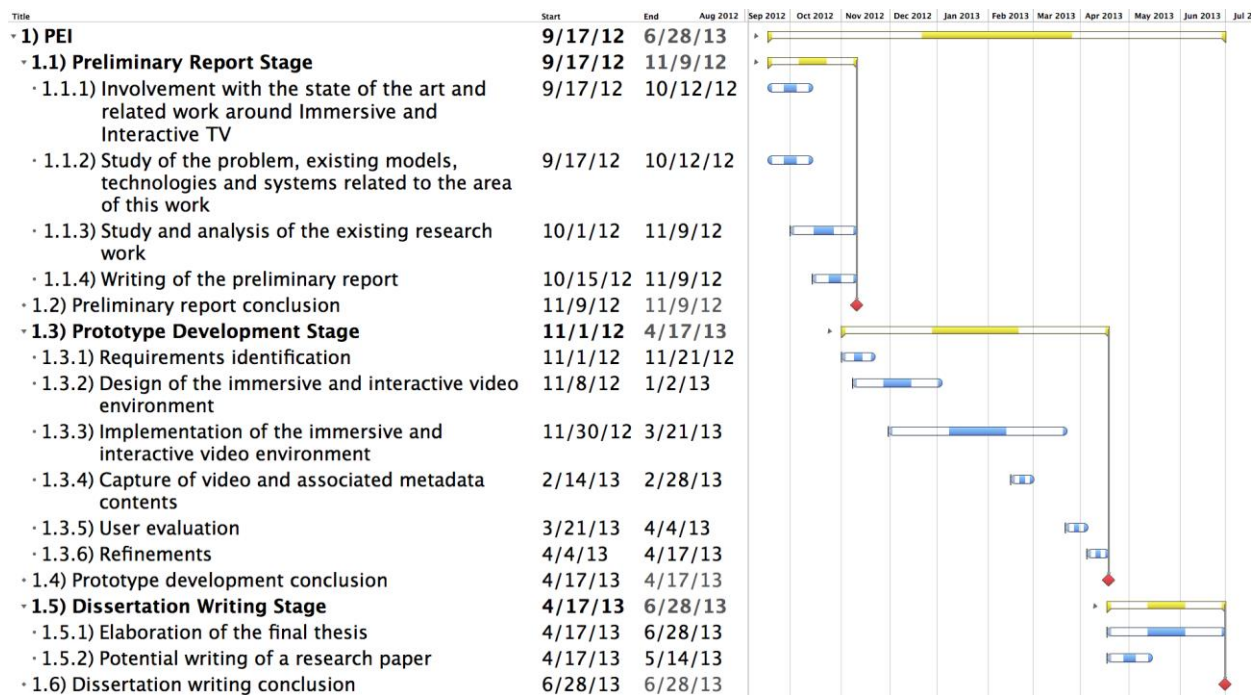


Figure 1.1: Initial Gantt Chart

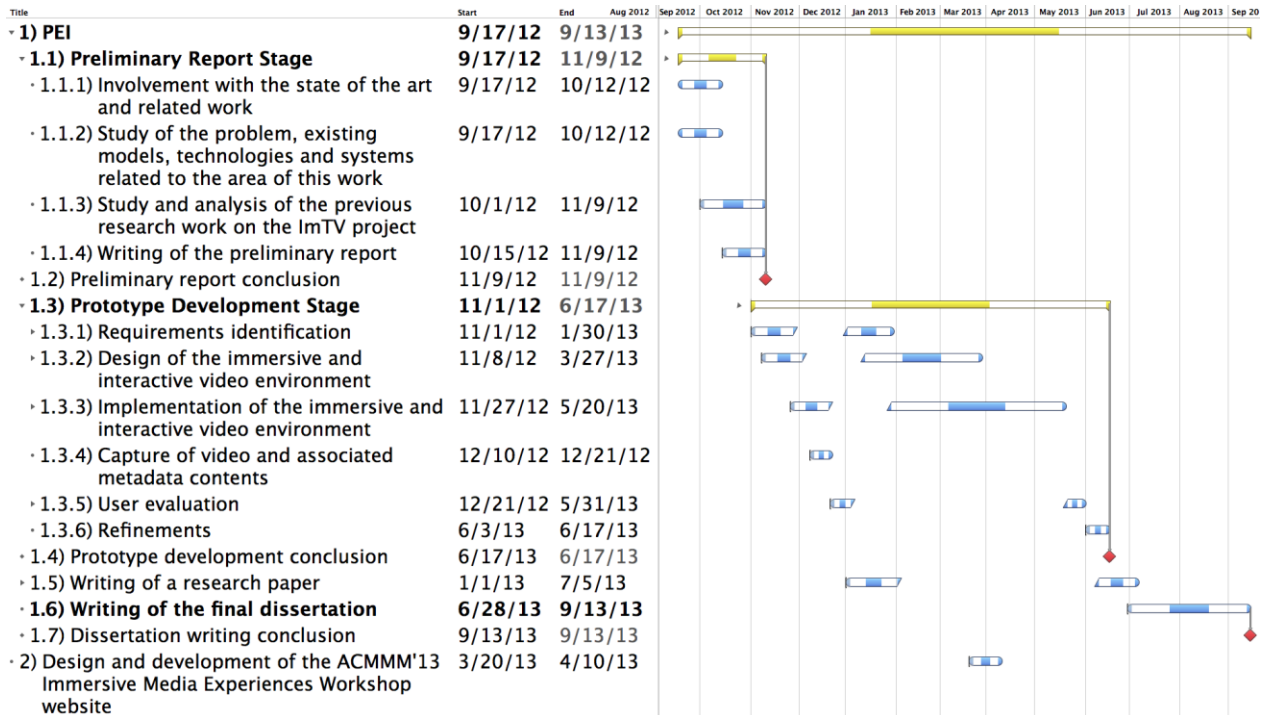


Figure 1.2: Final Gantt Chart

1.6 Document Structure

This document is structured around six chapters, as follows:

Chapter 1: Introduction – presents the main motivations and goals, outlining an overview of this thesis, summarising its research context, the contributions and the development plan.

Chapter 2: Related Work – presents the current state of the art in the research areas more closely related to this thesis. At first, the historical context of immersion in video is introduced, as this topic is at the core of this thesis. Afterwards, several relevant areas and examples of related work are presented and discussed, being that the core concepts are introduced along the way.

Chapter 3: Windy Sight Surfers – presents the design of Windy Sight Surfers, focusing on the video and metadata capture and publishing, the designed tools to search videos, and to increase immersion during viewing, in the two main categories of Perceptual Sensing and Context Awareness, followed by the Emotional dimension, which relates to video cataloguing and access and to user engagement and satisfaction evaluation, and lastly, the interaction with TV and wider screens.

Chapter 4: System Implementation – focuses on the implementation aspects of the features described in Chapter 3. Due to the exploratory nature of this thesis, several challenges emerged during implementation, which relate to both the functionalities

design (conceptual challenges), and to their implementation (implementation challenges). As such, these challenges are analysed and addressed in this chapter.

Chapter 5: User Evaluation – describes the user evaluation of the Windy Sight Surfers application, whose main objective was to investigate whether and in which conditions the designed features contribute to a more immersive environment for the user, according to the research questions established in section 1.2.

Chapter 6: Conclusions – summarizes and analyses the contributions of the accomplished work regarding its importance to the development of immersive video environments, and outlines directions towards future developments.

Chapter 2

State of the Art

This chapter presents the current state of the art in the research areas more closely related to this thesis. In order to do so, a first section introduces the main concepts related to Immersion, which is followed by a section that introduces the historical context of immersion in video, as this topic is at the core of this thesis. After this section, several relevant examples of related work are presented and discussed, being that the core concepts are introduced along the way. Advances in the Hypervideo and 360° Hypervideo fields are presented, followed by prior experiences related to Wind and 3D Audio are discussed, as these topics are closely related with some of the major components of this thesis. Later on, an Emotional Perspective is introduced, which motivates the analysis of several Emotional Models and Representations. In order to establish a link between Interactive TV and Mobile Environments, the concept of Social TV is introduced. Afterwards, the increasingly popular topic of Second Screen applications is presented, followed by the Guidance Systems in Maps and Recommendation Systems.

2.1 Immersion

Even if immersion is a wide concept, it can be assumed that immersion relates to an experience that creates the feeling in the users that they are part of the simulated environment (in some cases, as it occurs frequently with video, the environment is captured instead of simulated). However, given this concept's importance to the purpose of this work, immersion must firstly be defined carefully. A widely accepted definition of immersion follows:

“A stirring narrative in any medium can be experienced as a virtual reality because our brains are programmed to tune into stories with an intensity that can obliterate the world around us.... The experience of being transported to an elaborately simulated place is pleasurable in itself, regardless of the fantasy content. We refer to this experience as immersion.

Immersion is a metaphorical term derived from the physical experience of being submerged in water. We seek the same feeling from a psychologically immersive experience that we do from a plunge in the ocean or swimming pool: the sensation of being surrounded by a completely other reality, as different as water is from air, that takes over all of our attention, our whole perceptual apparatus.... in a participatory medium, immersion implies learning to swim, to do the things that the new environment makes possible.... the enjoyment of immersion as a participatory activity.”
(Murray, 1997)

Many scholars and scientists seem to agree that audio and photo-realism is not a necessary condition for an environment to produce in the user a sense of immersion. However, an aspect that can be taken for granted is that the more surrounding the virtual exhibition is (the bigger the screen, the more screens or the better the sound system) the more immersive it will be. Therefore, virtual reality producers have not changed their long-term focus from aiming for audio and photo-realism. It is anyway important to highlight that it is perfectly possible to create an extremely immersive environment in a standard computer, once immersion does not totally depend on the physical dimensions. According to McMahan, there are three conditions responsible for the production of an immersion sensation in an environment: (1) the user's expectations towards the environment must match the environment's conventions as much as possible; (2) the user's actions must have a non trivial impact in the environment; and (3) the environment conventions must be consistent, even if they do not meet the “real world” conventions (McMahan, 2003). It is therefore important to state that the user's psychological factor has a major role in the success or failure of the immersive environment. Lets consider the first movie of the Star Wars saga (1977): As soon as it debuted in cinemas it became a world phenomenon, mostly because of its realism. This happened because the technology used in the movie's production was the state of art at the time. As it was much more realistic than everything viewers had experienced to that day, the ones who saw the movie frequently felt completely immersed in the story. On the other hand, with technology's evolution, if someone goes to the cinema (this day) and watches a movie produced with the same technology George Lucas produced the first movie of the Star Wars saga, chances are the user will not feel as immersed by the created environment, even if the plot has all the requirements for a good story. This is mainly because the first condition has not been met, being that the user's expectations were not matched, as today's technology largely surpasses the technology used in the movie and users are expecting realism levels associated with today's technology.

Immersion can be divided in two main categories: Perceptual Immersion, were the user's physical senses are stimulated and the viewer has the feeling of being inside the

video; and Participative Immersion, where it is intended that users take a participatory role by contributing with their own content. According to Erkki Huhtamo (Huhtamo, 1995), immersion techniques can be split into two types: some techniques induce immersion as inward experiences, while others do it as outer experiences. As an example, chemical drugs clearly relate to inward experiences, while television, cinema or virtual games enable immersion to be experienced socially, as outer experiences. For the purposes of this work, the techniques considered are the ones that induce immersion as an outer experience.

With the goal of taking new immersive technologies to a growing customer base through in-situ demonstrations and experiments, Baker et al. (Baker et al, 2011) presented a combination of novel capture and display capabilities in delivering life-sized immersive 3D entertainment experiences. Although 3D technology is becoming consolidated by now, there are still some barriers to it. Producing 3D movies requires a lot of computer-generated information. Furthermore, unlike cinema, the broadcast of live events with 3D technology (with acceptable quality) continues to be a hard and expensive task. The reason for this is mainly related to the fact that, in a live environment, there is only one chance to acquire the desired video data (it is not possible to “rewind” real life events) and the production is not able to use the required computer-generated information, as it is a live environment. Baker addressed this problem, developing systems capable to generate good quality 3D (therefore creating immersive environments) and at the same time maintaining the production costs low. Unlike standard video formats, such as 16:9 HD, Baker considers much more panoramic formats in order to enhance the 3D experience (Figure 2.1). One such example is the 5:1 format, which can cover a much wider width. There are several cases where this might be useful, for example in a basketball match, with this technology it is possible to simultaneously film the whole width of the field, delivering an experience closer to the real one.



Figure 2.1: Representative image in the 5760x1080 format captured by one of the prototype cameras

Immersion and engagement have been studied more often in the context of games, with some properties shared with other environments. Brown e Cairns (Brown & Cairns, 2004) described work done to define immersion, being that their results are based on experiences of gamers. Gamers were interviewed about their gaming and

immersion experiences whilst gaming, being that the results suggest users experience different engagement levels. More specifically, Grounded Theory was used to construct a robust division of immersion into three levels: engagement, engrossment and total immersion.

Engagement represents the first (lowest) level of immersion and it must occur before any other level. The fewer barriers there are to enter this level, the less time, effort and attention are required to the gamer. Two initial barriers are the access (the gamer's preference and game controls) and the investment the gamer puts into the game. Once these barriers are transposed, the users begin to feel engaged. An engaged player is interested in the game and intends to continue playing. At this stage, what the experience lacks is the emotional level of attachment that is seen in later levels of immersion.

The next immersion level is **Engrossment**, which is achieved if the gamer transposes the game construction barrier. This barrier states that the game's characteristics are combined in such a way that the gamers' emotions are directly affected by the game. Due to the time, effort and attention put in the game, there is a high level of emotional investment at this immersion level. The player intends to continue playing and feels emotionally drained when he stops playing.

The final immersion level is **Total Immersion**, or presence. The barriers to this level are empathy and atmosphere. At this level, users feel disconnected from reality to such an extent that the game is all that matters. In other words, at this level of immersion, the game becomes the only thing that has an impact in the gamer's feelings and thoughts. Therefore, this study concludes that a shared concept of immersion exists, but instead of being a static experience, it is described as a scale of involvement with a game. The authors also state that immersion is not a necessary feature for enjoyment, once users choose the games based on mood.

Douglas and Hargadon (Douglas & Hargadon, 2000) used schema theory to define the characteristics of immersion and engagement in both conventional and new media. They examined how reader's experiences of these two different aesthetics may be enhanced or diminished by changes in interface design, options for navigation, and other features. After that, the authors studied "flow", a state where users are both immersed and engaged. The authors point out that, to this date, most studies related to the experience of reading hypertext documents have focused almost exclusively on readers' physical and cognitive encounters with texts, not on the affective pleasures readers derive from paging through mysteries, science fiction, and classic literature. Although not much has been done to this day, the affective dimension of interactive narratives can be readily explored by the use of schema theory to analyse hypertexts and

understand how interactive narratives frustrate or gratify readers in comparison with other texts. According to the authors, the pleasures of immersion stem from our being completely absorbed within the ebb and flow of a familiar narrative schema. The pleasures of engagement tend to come from our ability to recognize a work's overturning or conjoining conflicting schemas from a perspective outside the text, our perspective removed from any single schema. Concluding, the authors state that immersion and engagement are neither mutually exclusive properties nor polar opposites, despite the assumptions and assertions of most critics. Instead, most interactive texts necessarily rely on both, even though readers may perceive the narrative as entirely immersive or completely engaging.

As the video distribution over the Internet is becoming more and more common and its consumption switches from the TV screen to the computer, users' expectations are constantly increasing. In this context, Sekar et al. (Sekar et al, 2011) studied the impact of video quality in the user's engagement. For the purposes of the study, a set of different kinds of video was used: short video-on-demand (VoD), which corresponds to video clips with a length between 2 and 5 minutes; long VoD, which corresponds to video clips with a length between 35 and 60 minutes; and live content. Several video quality metrics were used in the evaluation, such as the join time, buffering ratio, average bitrate, rendering quality and rate of buffering events. The user's engagement was quantified, both at the "each video" and the "each user" levels. One of the main conclusions was that the time spent on buffering has the major impact on the user engagement across all types of content. Another conclusion was that the average bitrate has a much more important role in the live content than in the VoD (regarding the user's engagement).

2.2 Historical Context of Immersion in Video

This section aims to provide an insight through the historical context of video, and its ability in providing Immersive experiences to the viewers. Also, when discussing video, it is impossible to unbind this concept from the three main mediums that viewers use to access and consume video: Cinema, Television and, more recently, the Internet. For this reason, this section interlinks all these concepts, thus providing a description on how these three mediums changed the way video is used in the creation of Immersive experiences throughout the years.

When the 1880s first witnessed the invention of the movie camera, pioneers started to turn their attention to the then new film production industry. Namely, Eadweard Muybridge conducted an experiment that consisted of a series of photographs where the trotting of a horse was represented. Being developed under the motivation of providing

an answer to the question “are all four feet of a horse off the ground at the same time while trotting?”, this experiment would be considered one of the earliest silent films. The Lumière brothers made a great contribution by creating the first film production company and taking their productions to various venues in the world, such as London, New York and Bombay, which today represent some of the world’s movie capitals. Also Georges Méliès ([url-Melies](#)), contributed greatly to cinema by introducing several innovations regarding the use of special effects and the exploration of longer films, as at the time most films were still behind the minute long mark. With his advances, Méliès achieved great popularity, and therefore influenced other filmmakers, such as Charles Pathé, who would become one of the most dominant international movie producers.

As the popularity of films increased, movie theatres became more common. However, they were considered a cheaper, simpler way to provide entertainment when compared to the traditional theatre. The Nickelodeon was the first successful movie theatre, which introduced the current movie theatre paradigm (a set of contents composed a programme which would be changed weekly). By bringing the movie experience to the masses and turning it much more affordable than previously, the popularity of this concept led to an impressively fast expansion of the company to the point where it had approximately eighth thousand movie theatres around the United States in 1908. Movie theatres were now seen as a more convenient entertainment activity, as before the concept of movie theatre arrived, people would have to travel long distances to see the major theatre events.

In this context, during the beginning of the twentieth century, cinema in movie theatres was the main mean people had to access video as film. Therefore, giving the dimensions and comfort conditions of movie theatres cinema represented the first means to provide immersion through video. Although the popularity of movie theatres was increasing consistently, the Television was a technology already being investigated, since it first appeared in 1884 through Paul Nipkow’s work ([url-PaulNipkow](#)). However it was still a mean not available to the general public, and the dominance of the movie theatre paradigm remained until the 1930s, when the Television sets became affordable and started to be part of the furniture of families’ houses. In 1936, BBC carried out the first public TV broadcast in London. This event is considered a very important milestone and is regarded as the birth of television as we know it today. However, even though TV has spread to almost every house in the world, until recent times, TV was always associated with the producer-consumer paradigm, being the producer’s role represented by the TV broadcast companies while the viewers took part as the consumers. In other words, TV broadcasters chose to emit a certain schedule, while users were given no power to choose when to see what.

Despite the fact that the advent of Television did not give the user an active role in the contents selection, the spread of TVs through people's houses meant an important change in the way people consume video. TV was now the main mean to consume video, and this change resulted in a new idea of immersion through video: now users can view videos in the comfort of their homes, in a much more personalised environment.



Figure 2.2: Winky Dink and You

But, it was not until the 1950s that the first innovations towards a more interactive television started to emerge. Acclaimed as the first interactive TV show, “Winky Dink and You” (CBS) (Figure 2.2) presented the audiences with the new idea of enabling them to interact with the TV broadcasts. This interaction consisted of small films that were attached to the screen and then painted by the user, who would thus be able to participate in the story, for example by drawing a bridge that allowed Winky Dink to cross the river. Even though this is very far away from what is understood by “Interactive TV” nowadays, this event was crucial in the sense it created this entirely new concept. After that, several innovative approaches to iTV emerged. In 1959, the NBC’s “Today Show” (still on air) (url-TodayShow) introduced telephone calls in a TV broadcast for the first time. It is important to emphasize that this interaction approach remains highly popular in several program categories. For example, Hugo was a very popular TV show in Portugal (as well as in several other countries) during the 1990s, and consisted in an interactive TV show where users played a game using the telephone keys to control the character. Another milestone is the introduction of the World System Teletext (Department of Industry, 1989), which is still available to this day (in analog broadcasts) and presents a standard for coding and displaying teletext in television sets.

The transition to the twenty first century marked the emerging of technologies that enable the access to multiple Web tools through the TV. Namely, the “Video On Demand” (VOD) concept, which emerged during the nineties, was first launched as a commercial service in 1998 by Kingston ([url-KingstonVOD](#)) and was the first service to incorporate a TV broadcast and Internet connection in a single set-top box. Noticeably, Portugal pioneered the introduction of Digital Interactive Television (DiTV), when in 2001 the world’s first broadband interactive cable TV service was announced, resulting from a partnership between TV Cabo and Microsoft ([url-TVCaboDiTV](#)). This system introduced several innovative features, most notoriously allowing users to access both live and pre-recorded content in the set-top boxes, enabling access to e-commerce and e-banking information, providing iTV programs with access to additional content, a participatory role in several programs, and enabling the download of content to external storing devices (accessible in other devices anytime later). Interestingly, due to the market’s immaturity, in 2003 the company was forced to change the initial approach, by introducing a simplified system and gradually evolving with the introduction of new services.

Being that the empowerment of users by allowing them to select, watch and interact with video content on demand is the main characteristic of VOD systems, this technology completely revolutionised the way users consume video, now taking an eminently active role in the selection of the contents they desire to consume, resulting in new approach to immersive experiences.

Ever since, broadcasters have been offering increasingly powerful systems, focusing on the development of technologies that can enrich the user experience. Namely, most of the existent broadcasters deliver VOD systems through the Internet Protocol Television (IPTV), which is a system through which television services are delivered using the Internet protocol suite, and thus inherit several characteristics of the Internet network. Therefore, the tendency is for media convergence, where TVs will become more similar to computers, embodying several of their features. As an example, the Digital Video Broadcasting project developed the Multimedia Home Platform middleware system standard ([url-MHP](#)), which enables the reception over the broadcast channel (together with audio and video streams) and execution of interactive Java applications on a TV. Also, in 2006 Apple introduced the AppleTV ([url-AppleTV](#)), which uses a set-top box in order to connect several devices (smartphones or computers) to the TV and use its big screen to visualize multimedia content and browse the web through an appropriate interface. Furthermore, users can access content from services they subscribe on iTunes (Apple’s integrated multimedia platform), such as NetFlix or ESPN. Apple’s system uses a high speed wireless connection through which all the media content is streamed, instead of synched, which eliminates the need for the device

to contain storing capabilities, such as a hard drive, thus lowering its production costs. Another benefit of this approach is that it creates the possibility to stream content from every device that is also connected to the network. For example, users can stream a videogame from an iPad to the TV and view the content in the much bigger screen (Figure 2.3). Moreover, the company's more recent iCloud technology ([url-iCloud](http://www.apple.com/icloud/)), enables to stream content from the user's iCloud account directly to the AppleTV.



Figure 2.3: Content being streamed from an iPad to a TV through the AppleTV system

In parallel with the evolution of the TV environment, the Internet has also undergone through a remarkable evolution process, and the speed of this network has become far superior than it was during the 1990s. This led to the emergence of several new Internet technologies, which provide users with functionalities that would not be possible to deliver at the older speed rates. Namely, regarding video, a web application emerged in 2005 and would have great impact in the video industry – YouTube. It is a video-sharing website, on which users can upload, view and share videos, and it became widely popular shortly after its launching. Its traffic statistics increase each year, being that the 2013 data indicates one billion unique users visiting YouTube, over six billion hours of video watched per month and one hundred hours of video uploaded every minute ([url-YouTubeStatistics](http://www.youtube.com/t/ytstats)). This popularity resulted in a large amount of TV users shifting to the Internet in order to consume video, being one of the conclusions that can be drawn from these statistics the fact that users are very much interested in taking an active role on the content they consume, thus reinforcing the importance of interactive and participative TV.

The evolution of technology also let to the advances in cinema, and nowadays there are several approaches towards more immersive scenarios in the cinema industry. One

example is IMAX (url-IMAX), which is a motion picture film format and a set of cinema projection standards that enable to record and display images of far greater size and resolution than conventional film systems, thus offering a much more immersive viewing experience (Figure 2.4). Apart from the difference in the screen size (whose dimensions are up to 36 x 30 meters), this technology is able to reduce the distance between the viewers and the screen (as the resolution is much higher), and the screen is slightly angled, which gives users a sensation of being surrounded by the film.

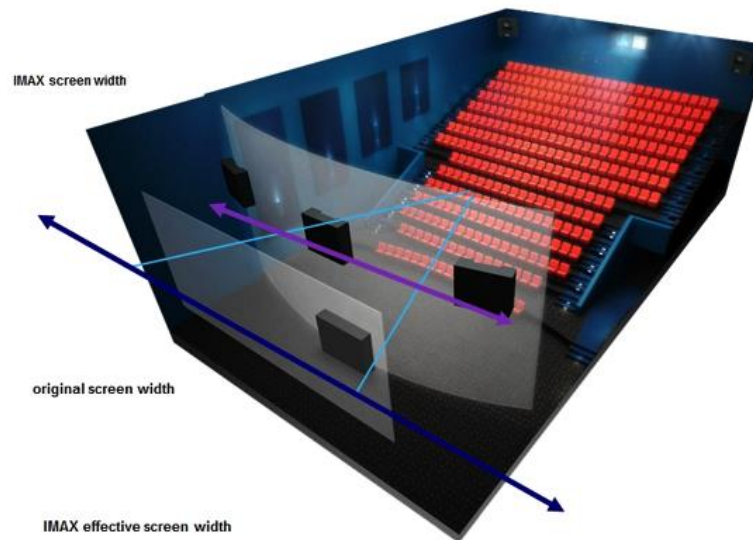


Figure 2.4: Comparison between IMAX and conventional film systems

In very recent times, research has been directing to the exploration of different senses in order to increase the immersive levels of a video experience. Three-dimensional (3D) film, which consists of a technique to produce and display videos with an enhanced illusion of depth perception, became very popular in recent years, albeit with moderate acceptance, due to concerns such as the requirement for additional devices (glasses), and the fact that prolonged viewing might cause eyestrain. However, even if 3D technology is still considered relatively new, several experiences have already been conducted regarding 4D Cinema, which improve the popular experiences provided in 3D Cinema with the exploration of other senses. Namely, a well-known experience related to the premiere of the movie *Avatar* in South Korea, where the movie experience was enhanced with sprinkling water, moving seats, smell of explosives and other features. This experience introduces the more recent notion of immersion: Multimodality. In this concept, several devices cooperate in order to increase the immersion levels. This concept is able to increase immersion through perceptual benefits of the exploitation of several senses (as opposed to only audio and video) or through the increased awareness of the users, given that they can access much more information on the contents they are consuming.

2.3 Hypervideo

Hypervideo refers to the video integration in hypermedia spaces. Therefore, in this context, video is not regarded as a mere illustration, but it can also be structured through links defined by spatial and temporal dimensions. More specifically, hypervideo may contain embedded links, which the user can follow, therefore making it possible to navigate within and across videos and between video and other hypermedia elements. That being said, hypervideo can be seen as hypermedia, where the dimension of time is central.

Even if the hypervideo concept dates from the early days of hypertext, when Ted Nelson extended his hypermedia model to include “branching movies” or “hyperfilms” (Nelson, 1964), hypervideo has not become as popular as the initial expectations predicted. This derives partially from technological constraints in hardware and supporting tools. Nonetheless, even if the progress has not been as much as initial thoughts predicted, big steps have been made since Ted Nelson’s first approach to hypervideo. For instance, in 1989 appears the hypermedia journal *Elastic* Charles (Brondmo & Davenport, 1989). This journal had an additional structural layer – the hyperlayer – over the video, sound and text material, which had already been edited and was in a “final” format. The aim of the structural layer - hyperlayer - was to link related portions of the journal's "stones" together to create the hypermedia environment.

Hirata introduced the Miyabi System in 1993 (Hirata et al, 1993), which provided the basis for the definition of dynamic links from content objects in video. Thus, the “media-based” browsing concept emerged. With the ability to automatically identify elements with certain characteristics in video, such as colour, form, structure, scenarios and sound, new opportunities to links were found and it became possible to follow objects along time.

HotVideo, which was presented in 1996 by IBM ([url-HotVideo](#)), is another example of hypervideo usage. In this application, the user was able to change to other resources by clicking dynamic objects in the video. In the following year MIT Media Lab’s project *Hyperssoap* (Dakss, 1998; [url-HyperSoap](#)) deepened this concept. *HyperSoap* was a soap opera in which the user could use an enhanced remote control to click on clothes, furniture and other content in order to obtain information on acquiring those products (Figure 2.5). This was a hypervideo example once specific objects could be selected by some kind of interface and the interactions with this objects changed the video flow.



Figure 2.5: Hypersoap

HyperCafe (Sawhney et al, 1996) was introduced in 1996 as an experimental prototype of hypervideo system that created “narrated video spaces”. The application placed the user in a virtual cafe, composed mainly of small digital video clips of actors chatting in conversations inside the café (Figure 2.6). This made it possible for the user to follow several conversations. Plus there were dynamic interaction opportunities through spatiotemporal, and textual links to present alternative narratives.



Figure 2.6: HyperCafe

Since HyperCafe was designed as a cinematic experience of hyper-linked video scenes, several of its design decisions were made in that direction: video was shown in black and white to produce a film-like grainy quality, the video sequences played out continuously and the user was unable to stop them at any point, as it happens in real life

situations. In other words, the user could simply navigate through the presented video and links, as these options enhanced the similarities of the application towards a real life visit to a cafe, where “real-time video” also plays out continuously. The used interface was a minimalistic one, with few visual artefacts in the screen, and the interaction and navigation was achieved through a mouse. All this made the application to create a more immersive ambience to the user, in the sense it closely resembled a real environment.

Girgensohn et al. (Girgensohn et al, 2003) developed an editor - Hyper-Hitchcock - and the “Detail-on-Demand” concept, which simplifies the process of editing interactive video by the use of more complex narratives that enable to present and edit interactive video at different detail levels: starting from the more resumed views to the more detailed ones. With this concept (Figure 2.7), the viewing and authoring interfaces are kept relatively simple, while supporting a great variety of interactive videos simultaneously. A direct manipulation environment enables users to combine videos and add hyperlinks between them. It also made possible to automatically generate a hypervideo composed of multiple video summary levels and navigational links between these summaries and the original video. This enabled users to interactively control several parameters, such as the amount of detail they intend to see, or the ability to swap to more detailed summaries (or to the original video).

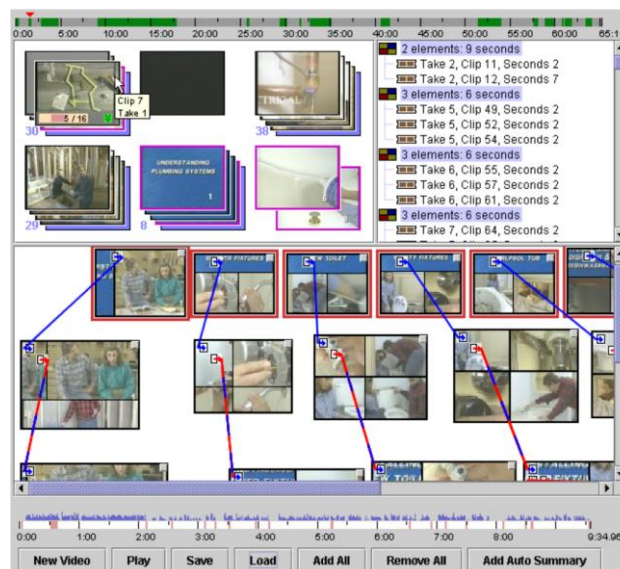


Figure 2.7: Hyper-Hitchcock

VideoClix (url-VideoClix) was launched in 2001 as a hypervideo authoring tool and today it stands among the most used SaaS (Software as a Service) solutions to distribute clickable video in the web and mobile devices. Using smartrack algorithms, VideoClix can identify people, places and products and create detailed hotspots for all objects in video. With this technology, viewers can interact with any object in

VideoClix Enabled content and therefore learn more about objects, purchase products, get storyline and bios, play along or receive general background info.

YouTube (url-YouTube) stands as one of the most popular video-sharing websites. It has not supported hypervideo for a long time but it started allowing users to add annotations to their videos (which work like hyperlinks) (url-YouTubeAnnotations).



Figure 2.8: YouTube Annotations

Figure 2.8 shows an example of a video with YouTube Annotations. Users can click these annotations, being the result the redirection to the related video or web page. This enables users to add interactive comments to their videos, which can lead them to add simple extra pieces of information, or to create stories with several possibilities, as the users choose the next scenes of the video.

With the goal of providing a principled basis for comparing systems as well as for developing interchange and interoperability standards, Halasz and Schwartz (Halasz & Schwartz, 1994) presented the Dexter Hypertext Reference Model, which is “*an attempt to capture, both formally and informally, the important abstractions found in a wide range of existing and future hypertext systems*”.

Furthermore, Bulterman et al. (Bulterman et al, 1991) presented the CMIF Multimedia Model, which is a model for representing and manipulating multimedia documents based on hierarchy structure. In the CMIF authoring environment different tasks of the authoring process are separated and the corresponding information is presented in three separate, but connected, views. The hierarchy view allows the author to define the structural relations between the media items that form the presentation. This structure is used to derive basic timing information, which is displayed in the channel view. The channel view shows the resource usage of the media items

composing the presentation, and enables to designate precise synchronization relations. The player, which represents the third view, is used to preview the presentation.

As an extension to the Dexter Hypertext Reference Model (Halasz & Schwartz, 1994) and the CMIF Multimedia Model (Bulterman et al, 1991), the Amsterdam Hypermedia Model (Hardman, 1998) introduced mechanisms that allowed the combination of “hyper-structured” information with dynamic multimedia information. Thus, the incorporation of time at a fundamental level in structured multimedia documents extended the hypertext notion of links to time-based media and compositions of different media.

The Synchronized Multimedia Integration Language (SMIL) (url-SMIL) is an XML based language with its roots in the Amsterdam Hypermedia Model and has been a part of W3C (url-W3C) recommendations since 1998. Until 2012, when HTML5 took over more attention as a way to address rich media on the web, SMIL enabled simple authoring of interactive audiovisual presentations through special players or browser plugins, assisting in the choreographing of multimedia presentations where video, audio, text and graphics are combined in real time.

As it has been previously referred, despite the advances in this field, hypervideo support has not been widely discussed and used, as it was expected to be. Therefore, there are several aspects of it that have not been addressed yet in a satisfactory manner, nor have been widely adopted. Namely, link awareness (where and when links exist) and high-level structuring constructs, which provide a richer and more contextualized integration of dynamic media. Furthermore, some of the more powerful existent hypervideo players are not yet integrated in the most common web browsers, and the current tendency relies on leaving to the authors/developers the responsibility of supporting these design decisions.

2.4 360° Hypervideo

Once it enables the presentation of a large amount and diversity of information in a rich cultural context through brief periods of time, video presents itself as one of the most effective ways to communicate. With 360° video, the central concept is the same as “regular” video: it consists of a sequence of still images representing scenes in motion. The difference lies on the fact that the angle of these images is much wider, namely 360° degrees. Therefore, it allows capturing and viewing the image all around, giving potential to the creation of environments where video can be viewed in a much more immersive way.

Video visualization is an activity where interaction can be very limited. Particularly, when one refers to web pages with incorporated video, the expectable

scenario is the simple inclusion of this type of content, followed by a linear visualization by the user. In such cases, the user's interaction with the video is generally limited to simple commands, such as play, pause, fast-forward or reverse. This led to the increased concern towards the improvement of video functionalities and the implementation of more interactive video interfaces in recent times. One of the most promising innovations is the possibility of implementing hypervideo over 360° video, thus providing true integration of video in hypermedia spaces. Once it provides flexible interaction mechanisms, this technology greatly improves video capabilities, by enabling the navigation through videos and their integration with deferent types of media. Furthermore, it opens the window to provide users with immersive experiences that recreate real scenarios captured in video.

However, despite the fact that hypervideo gives users greater control over the content, new challenges arise with the implementation of this technology in 360° video, mainly because a big portion of the video may be out of sight. Therefore, one of the big challenges is to develop effective 360° hypervideo players, so the users can understand the hypervideo structure and navigate effectively in a 360° hypervideo space.

Chambel et al. (Neng & Chambel, 2010 & 2011) in a previous work for the ImTV project identified potential challenges and benefits related to 360° hypervideo, presenting approaches to the design and development of immersive and interactive 360° hypervideo (Figure 2.9). Video immersion has a strong impact on the audience's emotions, in their feeling of presence and in their connection with the video; 360° videos have the ability to create highly immersive environments once they allow the feeling of being surrounded by the video; and Hypervideo stretches the boundaries even further, allowing users to interact with the video, explore it and navigate in a correlated information space. An interactive web application was designed and developed to explore the potencial and address the challenges of creating and accessing 360° hypervideos (Neng & Chambel, 2010 & 2011). This work linked two concepts that, by themselves create immersive ambiances: 360° video and hypervideo. The focus was established around the development of navigation mechanisms to help orientation and reduce cognitive load while viewing 360° videos. Through the web application, users are able pan around (left and right) continuously the 360° video, and several tools appear on the screen in order to support navigation and orientation. As an example, a minimap represents the 360° video's planar projection resized to be totally visible, with a red frame indicating (and allowing to change) the current angle of the 360° video being viewed (bottom of figure 2.9). By removing the viewing angle limitations, the minimap enhances the presence sensation while viewing 360° videos by improving the user's orientation through enhanced navigation mechanisms. Further developments on this work (Noronha et al, 2012; Noronha, 2012; Álvares, 2012) included the

synchronization between videos and maps, where the video trajectories and current position were highlighted on the map, and users may use the map to access other parts of current trajectories or other videos (Figure 2.10). Also, a new category of hypelinks was added to the system, which enable to reproduce a scene of a movie that was recorded on the same place as the current video, thereby giving the videos a temporal component, as their trajectories highlight videos from other occasions. Regarding video capture, a first approach was made towards the enrichment of video with additional information (metadata) in order to synchronize videos with maps. Therefore, it was developed a smartphone that, through the GPS, collected the geographical coordinates of the device periodically, while the camera was recording the video. Crossing this data with the video on post-processing enabled to synchronize videos and maps.



Figure 2.9: 360° Hypervideo Player

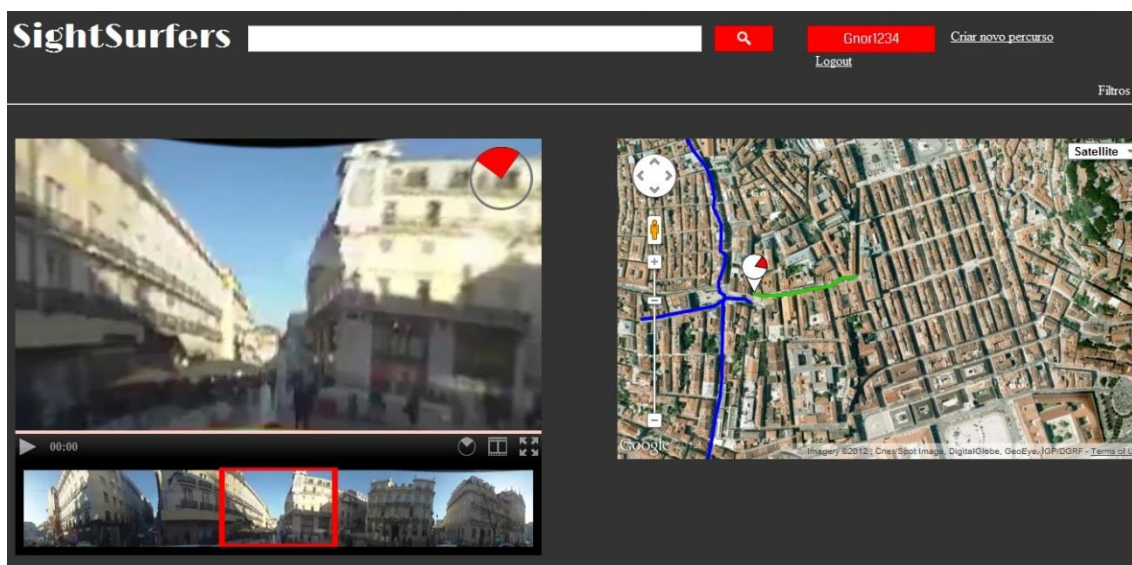


Figure 2.10: 360° Synchronization between 360° Hypervideo Player and maps

In a similar concept to 360° video, Omnidirectional Video (ODV) was defined as an innovative immersive medium that allows the spectator to be surrounded by video, which may be viewed with a head mounted display (HMD). Equipped with an orientation tracker this HMD shows a sub-image of the panoramic video that corresponds with the spectator's view direction and desired angle of view. Bleumers et al. (Bleumers et al, 2012) studied the problems and opportunities users anticipate about ODV. For instance, the study was guided by the following questions: “What characteristics make a TV program suited for enhancement with ODV according to adult digital TV viewers?” and “Do gamers and non-gamers have different ideas and preferences with regard to enhancing television through ODV?”. Answering the first question, the study’s results show that there are in fact certain genre-specific content elements that users claim to benefit from ODV. Examples include touristic programs, since they trigger the desire to explore. Besides these genre-specific contents, in general circumstances, when there is little progress in a program, viewers can use ODV as a temporary distraction. Regarding the second question, gamers especially appreciate the individual benefits of ODV, while the non-gamers tend to put the emphasis in the social benefits of ODV, being the final conclusion that both groups think of the ODV as something that they can benefit from. However, despite the benefits of ODV, it requires additional equipment (the head mounted display). Therefore, as the answer to the first question indicates that touristic programs benefit from ODV, there are grounds to further investigate whether it is possible to implement this concept in some platform that already exists and is commonly present in users’ daily life, and thus eliminating the resources associated to ODV (regarding the HMD), while keeping the results as approximate as possible. An example, which is related to the approach presented in this thesis, might be the use of mobile devices, such as tablets, as they are already present in users’ daily life and include a wide range of sensors, which eliminates the need for additional hardware equipment.

2.5 Wind Based Interfaces

Regarding the use of sensors and actuators, several interesting works have been accomplished. However, most of the examples target virtual reality instead of video, and do not target mobile environments, relying on heavy and very specific equipment. Furthermore, research into the effects of the wind (or air flow) has not been one of the central topics in the virtual reality community, even though, it was one of the senses included in the “Sensorama” system, one of the first immersive and multi-sensory (multimodal) reality systems (Heilig, 1962).

Regarding the use of sensors and actuators, Mendes (Mendes, 2010) presented technological art galleries where user’s movements influence video being displayed.

Motivation lies on the fact that interacting with the simulated environment enhances the users' experience and their feeling of becoming part of it, thus contributing to a feeling of belonging and strengthening the relationship with the environment. "B-wind!", one of the designed interactive installations, enables users to have the opportunity to perform an invisible character, the wind, being that the consequences of this character directly influenced the reproduced environment. Users are invisible, being that their physical presence is subtracted from the visual interface, but the result of their actions is presented in the real-time video through emphasised visual effects.

Moon and Kim (Moon & Kim, 2004) introduced a wind display system, called the "Wind Cube", which targeted virtual reality applications and included a number of fans attached to a cubical structure (Figure 2.11).

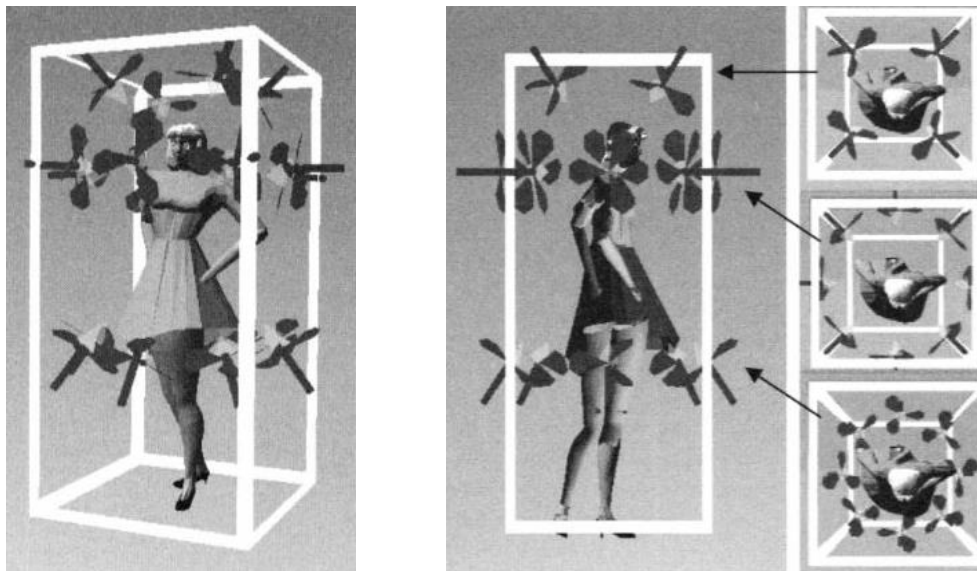


Figure 2.11: The WindCube. Small electric fans are attached to a frame in which the user stands

The authors carefully studied the design of the wind system, discussing the various issues that might occur when designing a wind system, such as the more adequate types of fans, the appropriated number of fans, their location and direction. Also, they developed an editor for the developer to manually define wind fields according to a virtual environment and a stationary user. The editor's main limitations were related to the fact that it was a two dimensional tool (not giving control to the developer over the third dimension). Also, although the editor allowed defining wind behaviour based on time, it did not provide any support for changing wind fields due to moving objects. When evaluating their wind system, Moon and Kim showed that the use of wind as output increased the sense of presence in a virtual reality environment. However, as in their system the movement occurred with a pre-defined animation path, their application did not allow any kind of user interaction.

Cardin et al. (Cardin et al, 2007) presented a haptic device designed to generate wind around the user's head. The system consists of a head mounted display driving 8 fan actuators regularly distributed around the device (Figura 2.12). The authors evaluated the device in the context of tele-operation through the use of a flight simulator application (also developed in this work's context) that uses the head mounted system to output wind direction and strength. Feedback shows improvements of the immersion levels when using the simulator paired with the head mounted display. Namely, users were expected to determine the wind direction, which they did with a variation of 8.5 degrees.



Figure 2.12: Head Mounted Display

While this system presents a plausible example of mobile wind display system, its evaluation has a strong limitation due to the fact that the user tests rely heavily on the flight simulator application, where there might be a big difference in the ability to pilot the airplane throughout the group of users that tested the system.

Kojima et al (Kojima et al, 2009) developed a wearable device for presenting the sensation of localised wind. The authors developed a prototype that applies wind around the user's ears, as the authors consider the ears are the most sensitive area to wind. Also, the authors manufactured a novel audio speaker specifically targeted at the display of wind. These small audio speakers based on a slit structure are able to apply a wind sensation more locally and with a shorter response time than small fans. When evaluating their prototype, Kojima stated that ears can perceive wind with high spatial resolution, and a local wind sensation can increase immersion.

Lehmann et al. (Lehmann et al, 2009) evaluated the feeling of presence when interacting with a 3D application, being that comparisons were made between a version

that used merely visual output, and two versions whose output were extended with stationary and head mounted wind systems. The results of the carried out experiment revealed a considerable increase in presence when using stationary and head mounted wind prototypes (Figure 2.13).

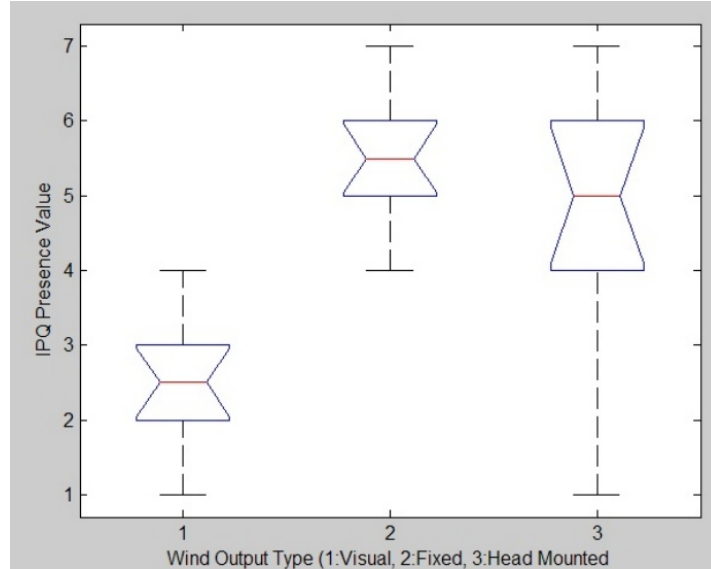


Figure 2.13: Lehmann notched box plot for presence in different wind prototypes

As it can be observed in Figure 2.13, results showed that wind output increases presence and indicated a tendency towards stationary wind output. However, this difference is not very significant, and therefore these results reinforce the idea that, although the maximum benefits are achieved with stationary wind systems, the final result is not significantly different from the values achieved with head mounted wind systems.

Analysing these works, most of these approaches target virtual reality environments - not video, nor mobile environments. Also, they tend to rely on heavy and very specific equipment, which decreases their feasibility, especially in mobile environments. Furthermore, none of them presents methods to capture wind metadata and couple it with the end result (e.g. video) as a way to increase realism and immersion.

2.6 3D Audio

Although most virtual reality and augmented reality applications have not focused on 3D-sound, this technology can significantly enhance realism and immersion, by trying to create a natural acoustic image of spatial sound sources within an artificial environment. Furthermore, most of these approaches address the production of film soundtracks or tend to address virtual reality scenarios (Dobler & Stampfl, 2004).

In what respects 3D sound engines, several options are available, featuring different technologies. Namely, some of the most powerful include Microsoft's DirectSound, OpenAL, Java's 3D Sound API, and Web Audio API. DirectSound ([url-DirectSound](#)), is a 3D sound engine, which is part of Microsoft's DirectX SDK ([url-DirectX](#)) and is programmable in C++. It is limited to Windows platforms, and it has not received any update since DirectX 8 (currently in version 11.1). OpenAL ([url-OpenAL](#)) is a cross platform (Windows, MacOS and Linux) that features sound management with sources and channels (sources are assigned to different channels), and provides tools for an easy integration with OpenGL ([url-OpenGL](#)). Java 3D Sound API is Java's sound engine ([url-Java3DSound](#)), which features sophisticated software architecture (organized through an intelligent class structure and providing tools for easy implementation). However, this engine is limited to stereo output, which means it can deliver 3D sound exclusively thorough headphones (as they are the only reasonable solution to dispose two sound sources uniformly to all users). Web Audio API is a Javascript sound engine for processing and synthesizing audio in web applications (consequently being cross platform), which provides an interface for positioning audio in a 3D space ([url-WebAudioAPI](#)).

Wave field synthesis (WFS) is a spatial audio rendering technique, which creates virtual acoustic environments, through the production of "artificial" wave fronts synthesized by a large number of individually driven speakers. The Virtual Source is the virtual starting point from which wave fronts originate. One of the main advantages of WFS is that the localization of virtual sources is not dependent on the listener's position, contrary to other spatialization techniques, such as stereo or surround sound.

Iosono ([url-Iosono](#)) is an audio system based on WFS that can be seen as a simulation of plane waves according to Huygen's law. In this system, an algorithm uses a 3D audio sample of the scene in reproduction to generate the secondary sound waves, which are needed to recreate the audio sample in the particular room where the reproduction is taking place. That information is then used to control a very large number of speakers (300-400), thus generating the desired "audio hologram".

Despite the sound quality advantages of the WFS technology, it has several drawbacks. Namely, this technology is very sensitive to room acoustics, because as it simulates the acoustic characteristics of the recording space, the acoustics of that space must be suppressed. Also, this technology faces high expenses, due to the large quantities of expensive materials its implementation requires. In the specific case of Iosono, this technology is mainly designed for cinema usage, thus having a limited number of applications.

Sound may be used to convey information related to movement, like speed and orientation, but for an intuitive mapping, it is necessary to take human perception into account. Providing users with a set of meaningful and intuitive control parameters is of great importance when designing sound synthesizers. This means that, in order to obtain synthesizers endowed with intuitive mapping abilities, human perception must be taken into account throughout the entire design process of the synthesizer. In practice, mapping between the basic signal parameters and the intuitive control device can be achieved by defining three layers: high-level parameters, which describe the way sounds are perceived through words, drawings, or gestures; middle-level parameters, which relate to the characteristics of the signal; and low-level parameters, which consist of the synthesis parameters. Merer et al. (Merer et al, 2013) proposed to characterize the general concept of motion evoked by sounds for synthesis and control purposes. When discussing the concept of motion evoked by sounds, the most obvious example of an association between motion and sound is the situation where a sound source is physical moving. However, certain sounds can also evoke a motion sensation metaphorically, as it happens frequently in music and in cartoon production. Therefore, designing an intuitive control strategy for motion evoked by sounds is still a great challenge for both music composers and sound designers. Focusing on monophonic sounds equally presented at both ears (and thus dismissing spatial cues), the authors began by determining how motion evoked by sounds can be described and characterized from a perceptual point of view, and then moved on to the identification of the high-level parameters involved in the mapping strategy. Using abstract sounds (for which the physical sources cannot be easily recognized), they asked users to describe how they perceived the motions evoked by these sounds by answering questionnaires and drawing trajectories, focusing on aspects like shape, direction, size, and speed. Thereafter, the authors conducted a listening test based in synthesized sounds combined with visual trajectories, which was performed to validate the relevance of the contribution of some variables. Upon the validation test, Merer draw up a typology of motion evoked by sounds based on the outcomes of this particular study. Providing an example of implementation of the proposed typology, the authors designed a generic sonification and control tool for evoking motion with sound from any input device generating continuous trajectories.

Nazime and Gromala (Nazime & Gromala, 2012) addressed sound mapping in virtual environments, with a concern on affective sound design to improve the level of immersion. Addressing the changes in the technologies for the application of sound in virtual environments, which are becoming adaptive and generative, they give sound a dynamic quality through controls that enable to edit the tonal characteristics in real-time. The authors use the concept behind Barry Truax's acoustic communication model

to consider two problems: how sound mediates information, and its importance to the listener through cognitive processing. Starting from this point, Nazime presented a model, addressing the use of procedural sound design techniques to enhance the communicative and pragmatic role of sound in virtual environments, which results in an environment where the listeners' experience is improved by engaging users to sounds (relating sounds to a specific time and space).

Manuel et al. (Manuel et al, 2012) investigated immersion in games through motion control and stereo audio reproduction. The authors developed a game system that tracks user motion around a room and changes the sound field to give users the sense of being in that virtual location. All sounds are programmed to change dynamically in audio level and reverberation level depending on the user's position in the room, and the distance from sources in the game environment. As the authors' objective was to test the immersive effects of audio simulation, the visual elements in the system were reduced to the required minimum visual aids. Through the conducted evaluation, Manuel concluded that soundfield interaction through motion tracking is capable of producing more enjoyable and immersive experiences than a handheld analogue controller, confirming that motion, along with high quality audio, can help create immersive states.

2.7 Emotions

It is hard to give a concrete definition of Emotion, mainly due to the wide variety of definitions that have been proposed (Kleinginna & Kleinginna, 1981). However, there is a consensus around the fact that emotions result from the response to events that are relevant to the individual's needs, objectives or concerns; and also that emotions are related with psychological, affective, behavioural and cognitive components (Brave & Nass, 2002).

2.7.1 Emotional Models

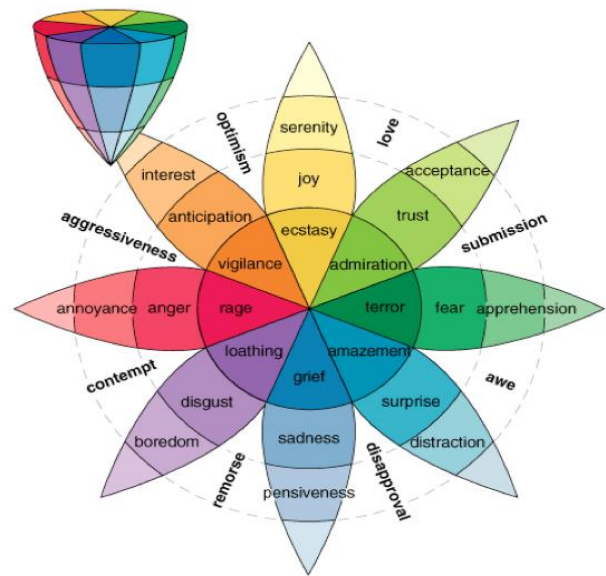
The main models relevant to our context relate to the Dimensional and Categorical Models.

Proposed by Russell, the **Dimensional Model** (Russell, 1980) argues that emotional states are triggered due to the cognitive interpretations of the main neural sensations. Russell advocates that, although facial expressions reveal as much representations of emotions as the rest of the human body (words, intonation, posture, movements), they provide primary information about the general emotional state of an individual. Russell proposed a model for emotions based on a two dimensional spatial circumflex, being that the horizontal axis represents valence, which highlights the

emotions' polarity (left side for negative emotions and right side for positive emotions); and the vertical axis represents arousal, which highlights the emotions' intensity (upper side for higher intensity and lower side for lower intensity). This model can be observed in Figure 2.14a.



a) Russell's Emotional Model



b) Plutchick Emotional Model

Figure 2.14: Emotional Models

The **Categorical Model** defines emotions as a number of discrete states that identify a certain behaviour and experience. Paul Ekman (Ekman, 1992) contributed greatly to this area by identifying six basic emotions based on facial expressions that are recognised across cultures: anger, disgust, fear, joy, sadness and surprise. A correspondence can be established between each of these emotions and its location on Russell's circumflex (Figure 2.14a).

Plutchick (Plutchik, 1980) advocates the existence of eight primary emotions (anger, fear, sadness, disgust, surprise, anticipation, trust, and joy), and used both categorical and dimensional models in order to define a 3D model (polarity, similarity, intensity) in a cone shape (Figure 2.14b). Emotions are represented around the center, in colours, and the vertical dimension highlights the emotions' intensity (which is reflected through the colours' tonality).

2.7.2 Emotional Representations and Visualizations

In the context of eliciting and visualising user emotions, there are some recent works. However, most of the examples don't focus on video.

Mappiness (url-Mappiness) maps happiness in UK, with the aim to better understand how people's feelings are affected by their current environment (air pollution, noise, and green spaces) and their own personal situation (friends, activities, places). Through the mobile phone, the application prompts users how they are feeling one or more times a day (Figure 2.15). Also, the system asks users whether they are alone or accompanied, the current location and the current activity. All this information is stored and crossed in order to understand the environment's influence in the user's state of mind.

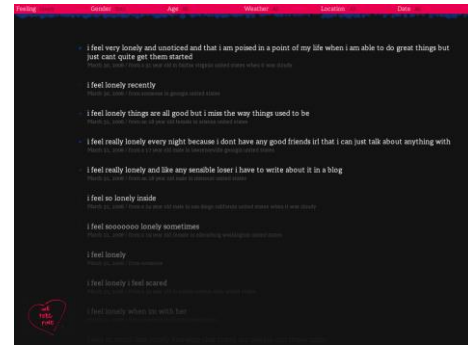


Figure 2.15: Mappiness

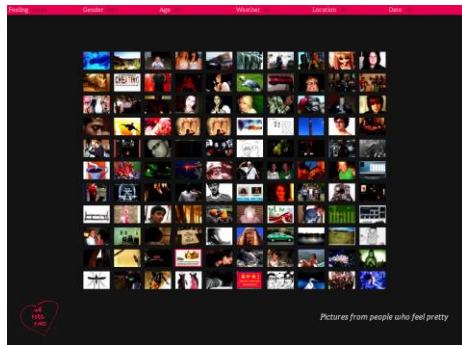
We Feel Fine (url-WeFeelFine) is a web application that harvests human feelings from a large number of weblogs, searching the occurrences of feelings after “I feel” and “I am feeling”. For each of the selected sentences, if additional information is available about the author and local weather conditions, these values are bundled with the sentence, thus being associated with the expressed feeling. The application offers an array of different search filters, such as feeling, gender, age, weather, location or date. There are six different options to view the search results:



a) Madness View



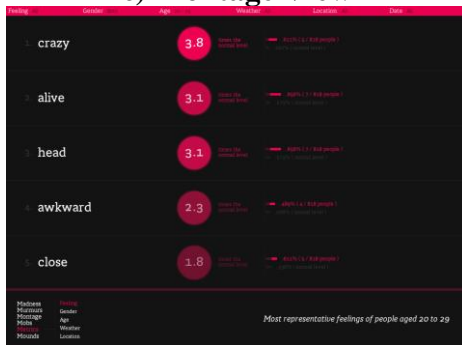
b) Murmurs View



c) Montage View



d) Subview Feelings from the Mobs View



e) Subview Feelings from the Metrics View



f) Mounds View

Figure 2.16: Wee Feel Fine's different views

1. Madness (Figure 2.16a), based on a physical particles system, where the particles move freely through the canvas, being that each particle represents feelings expressed by individuals, which are represented by a specified colour. Thus, particles are coloured according to the associated feelings.
2. Murmurs (Figure 2.16b), where particles float at the top of the canvas (right below the filtering menu), and one at a time descends into the list that highlights the contents of each particle.
3. Montage (Figure 2.16c), which consists of a grid holding images associated to selected sentences.
4. Mobs (Figure 2.16d), which defines five specific views (gender, feeling, weather, age and location), each of which self-organises the particles (colour, form, and order according to frequency).

5. Metrics (Figure 2.16e), defining the same five specific views as view 4, comparing the values associated to a sample with the mean of the entire set of information in the database.
6. Mounds (Figure 2.16f), this view presents all the existing feelings in the database, ordered by their frequency. Each feeling has a colour associated to it, which is represented through a bubble, whose size is relative to the number of occurrences.

Visch et al. (Visch et al, 2010) tested the effect of immersion on emotional responses and cognitive genre categorisation of film viewers, being that, during the experiment, films were presented in two conditions: in 3D (lower immersion) or CAVE (higher immersion). Viewers were asked to rate their own emotions and categorise the movies into four basic film genres (action, drama, comedy, and non-fiction). Regarding emotions, the experiment focused on the measurement of two distinct types of emotion:

- Fictional World emotions, which consist of emotions representing a response to the presented fictional events, such as sadness;
- Artefact emotions, which consist of emotions representing a response to the film as an artefact, such as fascination.

Results showed that higher immersion did not influence genre categorisation, but led to more intense emotions, both in the case of Fictional World and Artefact emotions. Furthermore, the authors state that the salience of emotional stimuli can be enhanced through higher immersion through parameters such as inclusion (e.g., attention filtering), extension (converging impact of impressions from various senses), surround conditions (sound effects), and vividness (a general condition of perceptual salience).

In a previous work in our research group (Oliveira et al, 2011 & 2013), videos were classified and accessed based on the emotions felt by users while watching them. This approach explored the use of three biosensors (respiration, heart rate, and galvanic skin response) to detect 5 of the 6 Ekman's basic emotions, and represent them with Plutchick's colors in the movies spaces and timelines in the interface.

Chambel et al. (Chambel et al, 2011), designed MovieClouds, an interactive web application for the access, exploration and visualization of movies based on the information conveyed in the different tracks or perspectives of its content, especially audio and subtitles, where most of the semantics is expressed, and with a special focus on the emotional dimensions expressed in the movies or felt by the viewers. For the overview, analysis, and exploratory browsing of both the movies collection and the individual movies, a tag cloud unifying-paradigm was adopted.

2.8 Social TV

The Social TV concept refers to a communicative and interactive TV environment and the technology that supports it. These systems often integrate voice communication, text chat, presence and context awareness, TV recommendations, ratings, or video-conferencing with the TV content, being that all these functionalities can be used directly on the screen or through auxiliary devices, such as smartphones (through specific applications). Being one of today's buzzwords, Social TV is becoming considered one of the main emerging technologies. Even if the idea of socialising around television is not a novelty, Social TV is paving the way to a cyber-living-room, and therefore providing a more interactive television (Jonietz, 2010). Following this thinking line can lead to the conclusion that Social TV can recapture the social aspects of television, which have been lost since the advent of multiple-screen households. This technology aims to connect viewers, even if they are not watching the same screen. It must be emphasized that Social TV, as a concept, is not bounded to some specific architecture by any means. In fact, Social TV is not necessarily limited to the television screen, being the computer screen and mobile devices examples of alternative architectures to deliver Social TV technology.

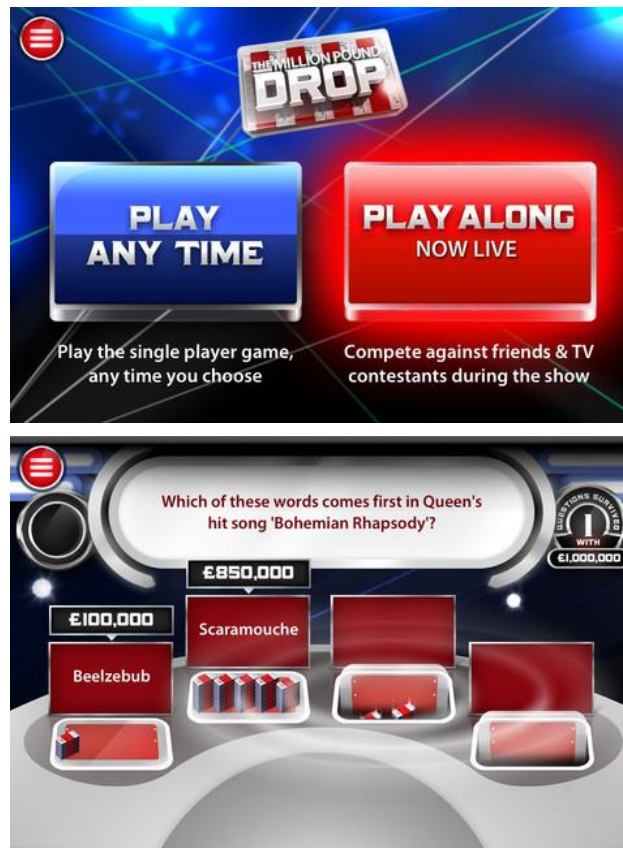


Figure 2.17: The Million Pound Drop's PlayAlong Application

In recent years, the Social TV has increasingly become more popular, being one of its benefits to the TV environment the ability to enhance the users' engagement by giving them the opportunity to become active participants on the TV system.

Visiware developed an application (Figure 2.17) that allows TV viewers to participate in a TV show while it is streaming over the Internet (in a synchronized way). The concept was called PlayAlong (url-PlayAlong). After that, several companies started to use this technology in their TV programs, being "Ten Million Pound Drop" (Endemol) one of them. In this program, TV viewers could play and answer quizzes as they would appear on TV.

Miso (url-Miso) is a Social TV based company that developed a second screen app (available for several popular mobile operating systems) (Figure 2.18) to enhance the TV viewing experience by introducing the TV Check-In concept, assuming itself as the TV's Foursquare. Foursquare technology enables the user to "check-in" a social network and share his location with whom he wants.

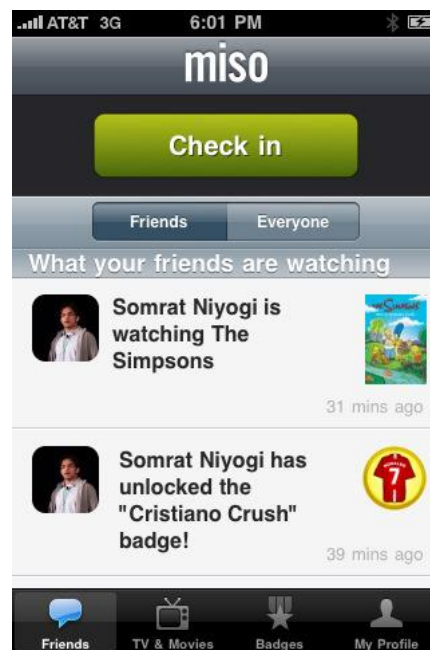


Figure 2.18: Miso

With the TV Check-In system users are able to check-in TV shows or movies in order to win points or trophies. Once checked-in, users can comment on episodes, they can "like" other user's comments, or they can rate their favourite shows, thus enabling other users to know the community's rating of the different TV shows available. Recently, the company released a mobile application called Sideshows that enables users (and other entities) to easily and effectively create additional contents about every TV show / movie, used as a second screen app. This technology presents several

characteristics that enhance the user's engagement because, with the TV Check-In concept, the user becomes an active participant on the respective TV system. The company's decision to partner with several TV chains brought new ways to "connect" the users (of those TV chains) with their favourite shows.

TVCheck (url-TVCheck) is an example of a social platform for mobile devices that applies Live Tv Image Recognition technology (Figure 2.19), enabling the user to check-in the social network just by pointing the smartphone camera to the TV screen. The channel and program are then recognised, thus enabling the users to share the programs they are watching with their friends via Facebook (or other popular social networks, such as Twitter), comment on the program, participate in live quizzes, win prizes or challenge friends. However, the app only identifies programmes running on the broadcast schedule, which presents a limitation, as any live changes to the broadcast schedule will not be dealt with, and extra contents the user might be viewing will not also be taken into account.



Figure 2.19: TVCheck

BuzMuzik (url-BuzMuzik), was recently introduced, and it is the first music channel integrating social media and enabling users to interact with it via Facebook, Tweeter or SMS. The channel has its own Facebook application (Figure 2.20), where users have total control over the music being played and can interact with the channel and other viewers (and with the TV channel).



Figure 2.20: Facebook’s BuzMuzik Application

Martins et al. (Martins et al, 2012) designed a Social TV chat system in the context of our ImTV project that integrates several methods, such as sentiment analysis, to measure TV viewer’s feedback with a multi-screen interaction paradigm. It detects the users’ state of mind towards a specific TV program by analysing their messages, which are shared with their friends through the system. This data is also used to present users with a graphical representation of the popularity of each show (Figure 2.21). One important aspect of this system is that, once the system follows the principle “one user / two screens”, the messages privacy is granted by separating the information presented in the shared screen from the information shown in the personal screen.

With the sentiment graphs, while users are zapping through the TV channels, they have instant access to the show’s popularity in the last minutes. Besides, these graphs have another use: they may be a much more accurate way to measure the real TV viewers preferences when compared to a simple audience measuring. However, there is a challenge in this approach, which is related to the fact that people might exchange messages that are unrelated to the programs being watched. This situation is particularly prone to happen when people loose interest in the program being watched.

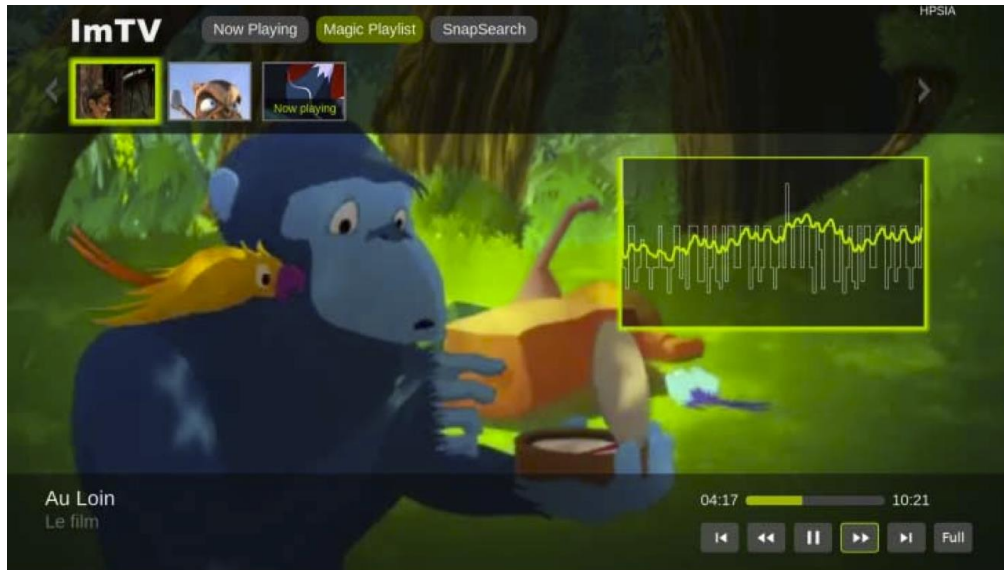


Figure 2.21: SentiTVchat with the sentiment graph

Wize (Almeida et al, 2012) is an application that combines some of the social games characteristics with some of the interactive TV characteristics, being the final objective to engage users through social TV games. The concept is based in the mix of two game formats, which have proven their quality: quiz and bet games. In order to achieve this combination, Wize adapted both these models, combined them with the TV consumer standards and enhanced them with social interaction dynamics, being the result a game that can be included in TV shows. By answering questions and making bets regarding TV shows, users collect points, which can be exchanged for prizes. Also, the game allows each user to define a list of friends, which may lead to increase the sense of belonging and visibility on the players network.

Avatar Theater is an experimental TV application that targets the enhancement of the shared viewing experience, by allowing users in separate locations to watch live or on demand programs synchronously. This application extends the traditional home viewing experience with a virtual theater, which virtually hosts other viewers who can share with each other their reactions to the content being watched. Using a multiplex movie theater as the interaction model, users can pick their personalized avatars, choose a theatre room, or create a new one. When interaction starts, viewers' avatars appear alongside each other, as if they were in a theater room (Figure 2.22). The real time communication features enable viewers to communicate using custom signals, text messages, sounds, interactions with the screen, and voice chat.



Figure 2.22: Avatar Theater (User avatars on the bottom and communication via emoticons on the left)

2.9 Second Screen

Second Screen refers to the use of additional electronic devices and applications in order to allow users to interact with the main source of output of the content they are consuming, such as TV shows, movies, music, or video games. Even more than that, these applications may extend the content provided by the main source of output. In recent times there has been a growing tendency of applying the second screen concept, which has provided technologies that create immersive environments, and enhance the immersive capabilities of existing ones, by surrounding the user with information of the contents they are watching. A common method to achieve this is relying on existing mobile devices, such as smartphones and tablets to interact with the main source of output and display extra contents. This concept is closely related to the TV environment, but that does not mean second screen apps are targeted specifically at the TV environment. There are several examples of second screen apps in multiple other contexts, such as video games, some of which are going to be described below. The relation between second screen and TV relates to the fact that Second Screen applications became popular mainly due to the advent of Social TV (described in section 2.8). The interaction principles behind Social TV stimulated the creation of new applications, and changed the content production principles. Therefore, Second Screen provides a parallel path of content and information with a multitude of applications in the environment. This section describes several works and technologies presenting innovations that can generate highly immersive environments.

2.9.1 TV Environments

Tseklevs et al. (Tseklevs et al, 2007) studied the use of second screens to mediate interaction with iTV services. The authors rely on previous work, where an initial study was carried out focusing on how participants use their iTV systems. Namely, the initial study revealed that public's perception related to security issues was stopping people from accepting this technology; users generally made limited use of the wide array of interactive features available (concentrating on the most familiar and essential functions); users frequently expressed concerns about hidden costs and security risks encountered when using iTV services (especially in the case of adults, where there was a generalized concern about their children entering competitions through iTV services without being aware of the costs involved); and users recurrently expressed the desire for rapid and direct access to content. Taking the results from the initial study into consideration, in this work the authors developed two prototypes (an on-screen prototype and a PDA prototype) that allowed controlling the TV and provided users with extra information (Figure 2.23). The evaluation of these prototypes concerned topics such as ease of use, multiuser compliance, speed and efficiency, concerns with cost, and parental control. The users' reception was successful, being the main advantages (according to users) related to the ease of use, efficiency, and the parental control (the portability of the mobile application allowed parents to control what their children were watching from another room in the house).



Figure 2.23: PDA prototype

IPTV is a relatively novel technology that delivers Digital TV through a packet switched network. Therefore, instead of using the traditional TV formats, IPTV uses Internet Protocols. This kind of TV network eased the introduction of second screen apps in the TV environment because, as it is based on the same principles as the Internet, communication with other devices in the Internet is facilitated. According to

the Alliance for Telecommunications Industry Solutions (url-ATIS), IPTV can be described as follows:

“IPTV is defined as the secure and reliable delivery to subscribers of entertainment video and related services. These services may include, for example, Live TV, Video On Demand (VOD) and Interactive TV (iTV). These services are delivered across an access agnostic, packet switched network that employs the IP protocol to transport the audio, video and control signals. In contrast to video over the public Internet, with IPTV deployments, network security and performance are tightly managed to ensure a superior entertainment experience, resulting in a compelling business environment for content providers, advertisers and customers alike.” (Alliance for Telecommunications Industry Solutions, 2005)

Bernhaupt et al. (Bernhaupt et al, 2012) recently presented a set of recommendations for the control of IPTV-Systems via smart phones based on the understanding of users practices and needs. One of the main conclusions of this study is that users prefer interactive systems where the TV is the interface and the mobile device is used mainly to control the content appearing on the TV. Another important recommendation states that the additional information of the TV channel currently selected must be presented on the borders of the TV screen, contrary to the majority of current IPTV user interfaces, which use the whole screen. On top of that, the screen space assigned to a piece of information must grow as that information becomes more important. This confirms industry attempts to design user interfaces that follow the users' engagement level. In the presented list appear several recommendations that most of current applications do not apply: the application must allow the control of every device related to the IPTV experience; design for usage-oriented scenarios; design for personalisation and personal use; support the user to control the connected home; enhance the overall user experience; and support touch and speech as interaction techniques.

Lochrie et al. (Lochrie et al, 2012) also studied the smartphones role in the TV as second screens. This work analysed tweets from the popular British TV show “The X Factor”, and compared them with tweets from TV shows in other formats. The results shown the richness of the information extracted in real time, and also the way audiences create their own parallel narratives of the TV show on tweeter. Results show that, in the iTV context, 40% of the tweets are submitted via mobile devices, which highlights the importance of these devices in iTV. These results are corroborated by a report from the International Data Corporation (IDC) (url-IDC), which highlights that, in 2010's fourth quarter, smartphones producers have for the first time surpassed the PCs producers in

terms of sales (number of units). As a conclusion, the authors state that smartphones are already becoming second screens for the TV, but not like broadcast services or personalised services. Instead, the audiences are creating their own forums to interact. Hence, the authors state it is imperative that TV companies understand the nature of this interaction and understand how it can be used with innovative interactive TV.

In a related work of our team in the context of the ImTV project, Prata et al. (Prata et al, 2011), addressed the effective design of video-based crossmedia services and interfaces with a particular emphasis on iTV, having developed a case study application – eiTV. The main focus of this approach is to provide the user with additional information regarding the TV content being viewed. Through the crossmedia concept, where several devices come into play (computers, iTVs, mobile devices), it is possible to generate a more flexible environment. The eiTV application takes cognitive and affective aspects into account in order to support interaction in its several cognition and interest modes. In order to give an answer to the learning opportunities created while watching video, the eiTV is capable of creating crossmedia personalized additional information in the form of web content that can be accessed from different devices in informal learning contexts.

An example of the use of audio information as a base technology of a social TV application is WiO (url-WiO). This mobile application “listens” and identifies the audio content streaming from the TV. Then it presents related information to the user. Information might include deals or discounts related with the products that may be being advertised on the TV. Despite the benefits this technology might introduce, it still has a major drawback: for the time being, before any content becomes able to be identified, it has to be “WiO-enabled”, which means an identifier has to be associated with it, and this can only be done by the service providers (both the content provider and the WiO responsible).

2.9.2 Video Game Environments

Recently, the video game industry started to show great interest in second screens. Video game consoles presented different approaches, as some include advanced controllers with second screen capabilities, while others announced applications for mobile devices platforms, such as iOS and Android. These different approaches have different benefits and drawbacks, and to illustrate them, some examples are presented next.

In 2011, Nintendo announced its eighth generation video game console – the Wii U (url-WiiU) (Figure 2.24). This is the successor of the Nintendo Wii, which was the bestseller console from the seventh generation of video game consoles. One of the main

characteristics of the Wii U (which consists in one of the main novelties in video game consoles design) is that it has two kinds of controllers: a conventional one, and the Wii U GamePad, which is an advanced controller that features a touch screen. This controller can function as a second screen, where it is used to expand the contents being displayed in the main screen, as can be observed in figure 2.21. However, it may be used as the only screen of the console, thus eliminating the mandatory request for a TV set to play.



Figure 2.24: Wii U GamePad

Besides the traditional input methods, and inheriting lots of characteristics from the Tablet market, this controller has a built in accelerometer, a gyroscope, a camera, and NFC (Near Field Communication) technology. This hardware creates a new kind of experience: asymmetric competition. In other words, the user with the GamePad has a certain experience and is able to win in a certain way, while the other players have other ways to achieve victory. This is derived from the fact that the GamePad enables a new array of entertainment opportunities and, therefore, the player with the GamePad can get the responsibility to achieve a certain task, which is only possible to achieve with the new technology, while the other players have other tasks to achieve in the game. As outlined, this approach introduces great benefits to the video game console environment. However, it relies on specific hardware, which means added costs to the end user. Furthermore, this additional hardware fits just the purpose of this application. Next, a different approach is presented, and it is discussed how it compares to the Wii U GamePad.

Microsoft introduced Xbox SmartGlass in 2012 (url-XboxSmartGlass). This is a mobile application that complements their video game console Xbox 360 and is available for the major mobile platforms (Android, iOS and Windows Phone). This

application provides users with features like interactive companion guides, behind-the-scenes commentary, and real-time game strategy (Figure 2.25). Furthermore, users can use their smartphone or tablet as a controller to surf the web in the TV (through the Xbox console). This approach might lack some of the features the Wii U GamePad provides. For instance, it is not able to replace the TV screen as the main screen of the console. On the other side, integrating the most pervasive mobile platforms with the video game console environment means connecting it to an immensely vast information network, which enables a much more scalable system, being that multiple entities can develop for mobile platforms, and in this context, contribute to the Xbox environment.



Figure 2.25: Xbox SmartGlass

2.9.3 Other Environments

Until now, the presented examples concern big and relatively uniform environments, such as TV networks and video game consoles. However, second screens are being introduced even in much more specific environments. Increasingly, more products are launched to the market with an associated second screen application. One such example is the wearable video camera GoPro. GoPro recently launched a mobile application for smartphones and tablets that enables to remotely control one or more GoPro cameras simultaneously (Figure 2.26). It is possible to configure every control of the camera. Besides, the cameras have the ability to stream the video being recorded in real time, so the user can watch the video being recorded by all the cameras in a live environment. Additionally, users can share online contents stored in the camera's SD card through the application, which means that videos are transferred from the video camera to the mobile device, which uploads them to the desired online network.

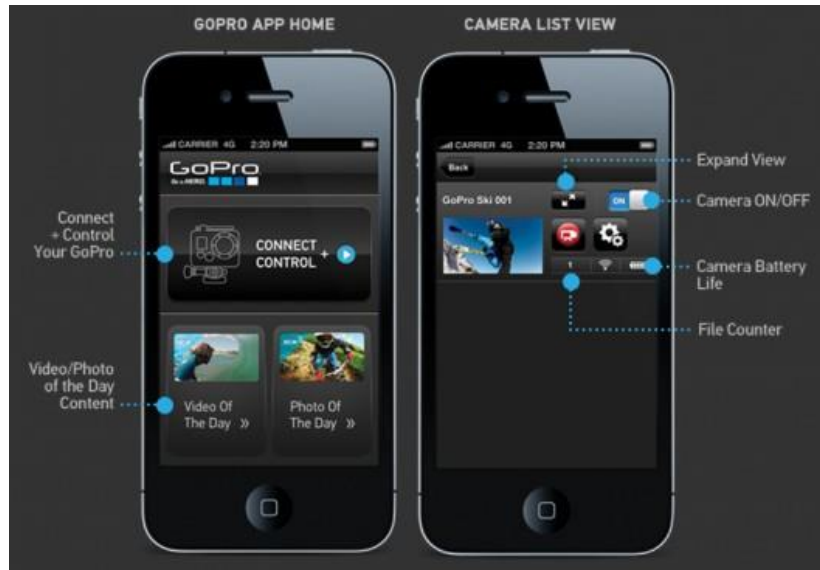


Figure 2.26: GoPro's Mobile App

On its behalf, the music industry has been embracing the use of second screen apps through mobile platforms in order to extend the capabilities of musical instruments. This approach not only extends control over the hardware, but some applications even add new features to the hardware components.

Akai introduced the SynthStation keyboard controller line. A keyboard controller is similar to a digital piano, but it does not include any soundboard. It consists just of the hardware equipment and therefore needs to be connected to some kind of equipment in order to produce sound (usually a computer or sound module).



Figure 2.27: Akai's SynthStation keyboard controller

The SynthStation line, contains a companion second screen application, whose capabilities include extended control of the keyboard beyond the hardware controls

(url-AkaiSynthStation). But more importantly, the application includes several musical instruments and other features, such as a sequencer (Figure 2.27). With these features, this second screen application transforms a keyboard controller in a fully featured musical instrument. Furthermore, Akai released the SynthStation SDK, which allows third-party developers to design and create applications for Akai's keyboard controllers.

Line 6 recently introduced the StageScape M20d live sound mixing system (url-Line6StageScape), a fully featured digital mixing system designed with ease-of-use in mind. Rather than the classic mixing board interface, which consists of hundreds of knobs and faders, the StageScape presents an innovative touchscreen interface (Figure 2.28a). Furthermore, the StageScape provides a second screen application, which enables sound engineers to be able to leave the mixing desk area and continue to have total control over the sound system from other places in the concert hall (Figure 2.28b). This is of great importance, as sound engineers can adjust the sound characteristics in each area of the concert hall according to the sound they really listen and analyse, and thus improving the experience of the audience.



a) StageScape sound mixing system

b) StageScape's second screen application

Figure 2.28: Line6 StageScape M20d

2.10 Maps and Georeferenced Guidance

Mapping is the process of designing, implementing, generating and delivering maps under a certain platform, and nowadays, there are several mapping systems that allow users to navigate maps. Most of these systems use Web Mapping, which consists in the described mapping process, delivered in the World Wide Web as a product. The use of the web as a dissemination medium for maps had a big impact in society, and is regarded as a major advancement in cartography. The first approaches towards web

mapping systems were primarily static, but this technology has evolved greatly, and today's web maps are fully interactive and integrate multiple media.

Google Maps (url-GoogleMaps) is most likely the best-known web mapping system, and serves as an engine that powers most of the map-based services available, including Google's own map related technologies. Based on the Mercator Projection, Google Maps include a wide array of interactive features that associated information with geographical locations through hyperlinks, which are designated Markers and can contain information such as audio, video, image, or text content (Figure 2.29).

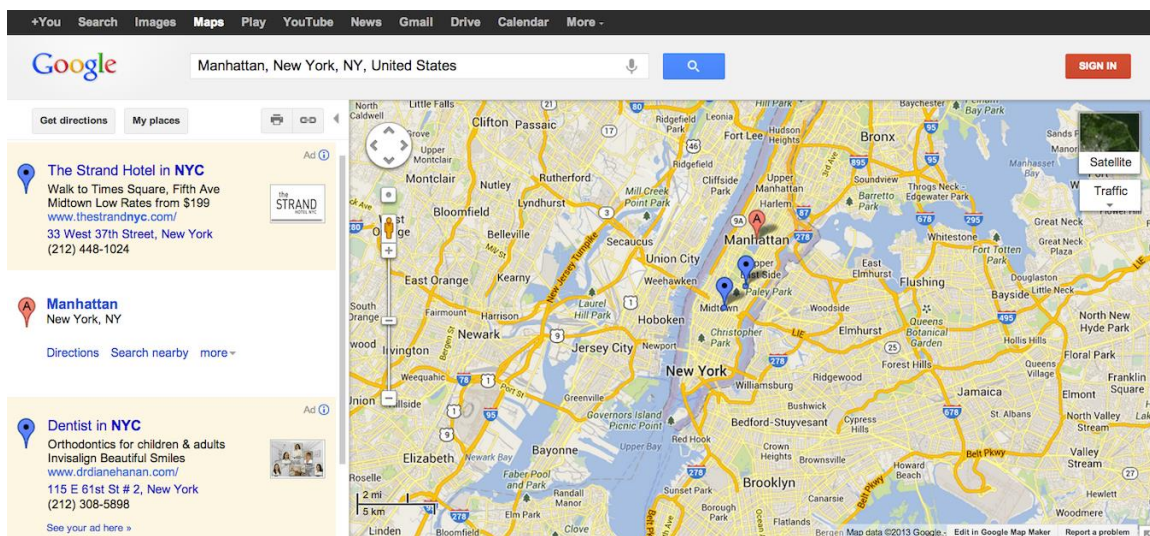


Figure 2.29: Google Maps Web Application



Figure 2.30: Google Maps Mobile Application

Initially intended as a browser technology for computers (and thus based in JavaScript), due to the improvement of mobile devices' capabilities, Google Maps is

supported by several mobile platforms (such as Android, iOS, Windows Mobile) in addition to the web browser application, being that in mobile devices that feature a GPS the Google Maps application allows to track the current location of the user (Figure 2.30).

Google allows developers to freely integrate Google Maps into their websites or mobile applications by providing an API for the multiple platforms it supports (url-GoogleMapsAPI). This API has paved the way for multiple developers to build upon this mapping system and enhance their applications with new functionalities, some of which will be described next.

Nike and Apple introduced the Nike+ running kit in 2006 (url-Nike+), which is an application targeted at measuring and recording the distance and pace of a walk or run. Initially, the Nike+ kit consisted of a small transmitter device that was attached to or embedded in a shoe and communicated with a receiver plugged into an iPod Nano or directly to the iPhone/iPod Touch through Bluetooth (being that the mobile device hosts the Nike+ application). However, a more recent version of the Nike+ application, which eliminated the need for the shoe sensor, since the newer versions of iPhones and iPods contain GPS and Accelerometer sensors. In 2012 the Nike+ application was launched to the Android platform.



Figure 2.31: Nike+ Application

Through the GPS, the application allows users to map their runs and track their progress, thus working as a motivational companion to runners. Several data is tracked, including georeferences, distance, pace, time and calories burned (Figure 2.31). Feedback is given to the user both after and during the run. After submitting the run to the server, users can log into their accounts in the web application and see their runs

including route and elevation. Furthermore, the application has a strong social component, as Nike+ strives to create a global running community. Several features contribute to this goal, such as NikeFuel, which is a measure that takes into account all the activities of a user's athletic life. Therefore, users compete globally to achieve higher NikeFuel values. Also, the application is integrated with social networks, allowing users to Post the start of a run to Facebook and hear real-time cheers for each like or comment they receive.

ATC9K (url-ATC9K) is an example of georeferenced video by integrating video end maps (Figure 2.32). The system consists of a video camera with an integrated GPS sensor that enables users to map their location, speed and distance travelled. The camera includes a software application (for computers) that enables users to view the recorded videos while tracking the movement through Google Maps.



Figure 2.32: ATC9K Application

2.11 Recommendation Systems

The Internet hosts an immensely vast set of content, whose size is constantly and rapidly increasing. In order to assist the cataloguing of content, and deal with the problems this task faces due to the amount of information, recommendations system have been gaining popularity. Using metadata (which provides a description of each content element) and the users' history (collected while users are interacting with applications), the recommendation systems outline each user's usage profile. With this information, applications can then predict users' interests, and thus present them with related contents.

The NoTube project (Aroyo et al, 2011; url-NoTube) used Semantic Web technologies as a tool to connect TV content and the Web through Linked Open Data.

The main contribution of NoTube relates to the specification of protocols and APIs to support a variety of realistic user scenarios for experiencing future TV. More specifically, the NoTube project explored the recommendation of TV content based on user information collected from as much web content as possible.

Semantic technologies relate to standards and technologies that allow a machine-understandable and machine-processable representation of the meaning of digital content, in order to enhance applications, by improving their intelligence, responsiveness and personalisation. The NoTube project uses these technologies to address semantic annotation (enabling new types of applications that have a better knowledge over what users want and what programmes describe) and user interests from the social web (extracting knowledge from user's TV viewing activity and social web presence, so that applications can implicitly determine users' interests, and match them to TV programmes).

Ferman et al. (Ferman et al, 2003) proposed algorithms for automatically determining a user's profile based on the content usage history (profiling agent), and automatically filtering content according to the user's profile (filtering agent). In the designed approach, the authors use a fuzzy inference system to construct and (periodically) update the user preferences. These preferences are then paired with structured metadata related to each of the consumed content entities to identify what types of content the user prefers the most. Describing the user profiling technique in more detail, it is important to highlight that the usage history description follows the MPEG-7 standard, and thus provides a compact collection of user actions, as well as identifiers of content associated with these actions. Also, given the fact that, initially, there is not a user profile available, the authors consider two processing modes for generating and updating user preference descriptions:

- *Batch mode*, used when no initial user profile is available – user preferences are constructed from the start using the entire usage history available;
- *Incremental mode*, used when a user profile is available – the information from the user profile itself is paired with the more recent entries in the usage history log, in order to incrementally update the preference description to reflect the changes in the viewing choices of the user.

This way, Ferman's approach provides a real time filtering mechanism that generates dynamic content lists in response to user requests.

Weiß et al. (Weiß et al, 2008) introduced an approach that used user profile information in order to personalize digital multimedia content. Arguing that the usual approach of content-based recommendation mechanisms, which perform an item-to-item based matching and match unknown content with other similar content that has

already been rated by the user, may cause wrong or misleading ratings, the authors propose an approach based on an item-to-profile matching mechanism, considering content entities with the same attributes, not just the most similar ones. In this context, the authors introduce mechanisms for the creation of user profiles. More specifically, two ways of profiling are considered: Explicit Profiling (where users explicitly state if and how much they liked or disliked content after consuming it); and Implicit Profiling (where ratings are automatically derived by analysing the user behaviour). Regarding the content-based recommendation algorithm, it estimates the users' interest in unknown content by matching their profile to metadata descriptions of the content, consisting of item-to-profile based matching mechanism, as it was already stated.

Chapter 3

Windy Sight Surfers

The main objective of this thesis was the design and deployment of techniques that may enhance the immersiveness of a 360° video environment.

When trying to increase immersion, two main categories of techniques arise. On one side stand the techniques that strive to increase immersion through giving the user a better context awareness of the environment in question. Perhaps one of the best examples of these techniques can be found on books, as really engaging novels, all the way from the classic authors to J. K. Rowling, are able to create highly immersive environments just with words, meaning their stories are described in such a way users get involved with it. The other category of techniques intends to enhance the immersive capabilities of an environment by introducing features, which explore user senses, and use that as a means to make the user feel inside the environment. As an example, 360° video is in itself a technology that delivers to the user a much more immersive video experience. For the purposes of this thesis, these categories are referred throughout this document as *Context Awareness* and *Perceptual Sensing*.

In this thesis, several techniques were designed, developed and tested, both from the Context Awareness and Perceptual Sensing categories. In order to do so, and building on previous work (Noronha, 2012; Álvares, 2012), Windy Sight Surfers, a 360° video application that comprises all the designed features, was created.

Mobile devices are proliferating and their popularity is increasing rapidly. Also, mobile devices provide a wide range of sensors and actuators, which is ever increasing and represents a strong potential to allow augmenting human-computer interaction and, more specifically, supporting more powerful and immersive video user experiences. Therefore, the Windy Sight Surfers application is primarily intended as a mobile application targeting the Android platform. Still, the work of this thesis also focuses on using mobile devices as second screens to interact with iTVs, which means there is also the possibility to use the application in this kind of environments.

This section presents our main design options so far. Firstly, the functional and non-functional requirements and the use cases of the referred application are presented. Afterwards, focus shifts to the Windy Sight Surfers application. This chapter's structure was based on the most natural sequence of interaction (user's capture, publish, search and view videos). Therefore, this chapter focuses especially on the video and metadata capture and publishing, the designed tools to search videos, the designed features to increase immersion during viewing, which divide into two main categories (Perceptual Sensing and Context Awareness), the Emotional dimension, which relates to video cataloguing and access and to the evaluation of user engagement and satisfaction and, lastly, the interaction with TV screens.

As the Windy Sight Surfers application comprises all the designed features, each of the different categories is described between sections 3.2 and 3.9.

3.1 Requirements Specification

This section intends to present a complete description of the system behaviour. Therefore, the system functional and non-functional requirements are specified. Also, as it focuses on the user interaction with the system, a set of use cases is presented.

3.1.1 Functional and Non-Functional Requirements

Functional Requirements that Windy Sight Surfers must meet:

- Provide functionalities to capture 360° video metadata (such as geo-coordinates, speed, or weather conditions);
- Provide a submission service compliant with 360° video and its associated metadata;
- Store the submitted videos and their associated metadata.
- Provide tools for editing metadata characteristics before submission;
- Provide tools for searching videos through a set of keywords and filters;
- Provide tools for searching videos through a map, by specifying areas or paths in which the user is interested;
- Display the videos being watched in real time in the map;
- Display search results in a list of videos;
- Display search results on the map;
- Allow to select a video from the map;
- Enable the visualization of 360° videos in a circular perspective around the viewer;
- Provide mechanisms to rotate the visualization angle of the 360° videos;

- Provide tools to navigate between videos in real time through hyperlinks;
- Provide link awareness through tools to alert the user about off-screen content, especially hyperlinks;
- Synchronize the video with the map, moving the video markers as the video plays and accessing video in the moments corresponding to the selected position on the map trajectories;
- Increase immersion levels by giving the user a more realistic sensation of movement – for e.g. through wind;
- Use three dimensional audio in order allow the user to easily identify the orientation of the video;
- Use three dimensional audio in order to increase immersion levels by giving the user a more realistic sensation of movement;
- Display an overlay during video reproduction with permanent information features, in order to increase immersion levels by providing the user with a set of context awareness information;
- Display an overlay during video reproduction with momentary information features, in order to increase immersion levels by providing the user with a set of context awareness information;
- Provide a video cataloguing system according to users' expressed emotions while viewing the videos;
- Display dominant emotions of the videos in each area of the map;
- Provide the users with statistical information about their personal emotional history related to the application;
- Provide video recommendations to each user, based on their personal emotional preferences;
- Synchronize a mobile application with a TV, thus using the mobile application as a second screen for controlling the content on the TV and display extra content, in order to provide a more immersive environment.

Non-Functional Requirements that Windy Sight Surfers must meet:

- Be fault-tolerant;
- Be fast and efficient;
- Be designed accordingly to the Modular Software design principles, which focuses on several quality concepts, such as changeability, reusability, comprehensibility, or testability;
- Be designed with usability and ease of use in mind.

3.1.2 Use Cases

This section aims to present the various ways in which users interact with the system. Therefore, a Use Case Diagram is first presented (Figure 3.1), depicting the main use cases. Afterwards, the textual Use Cases are presented, accompanied by the corresponding Sequence Diagrams, which highlight how processes operate with one another and in which order.

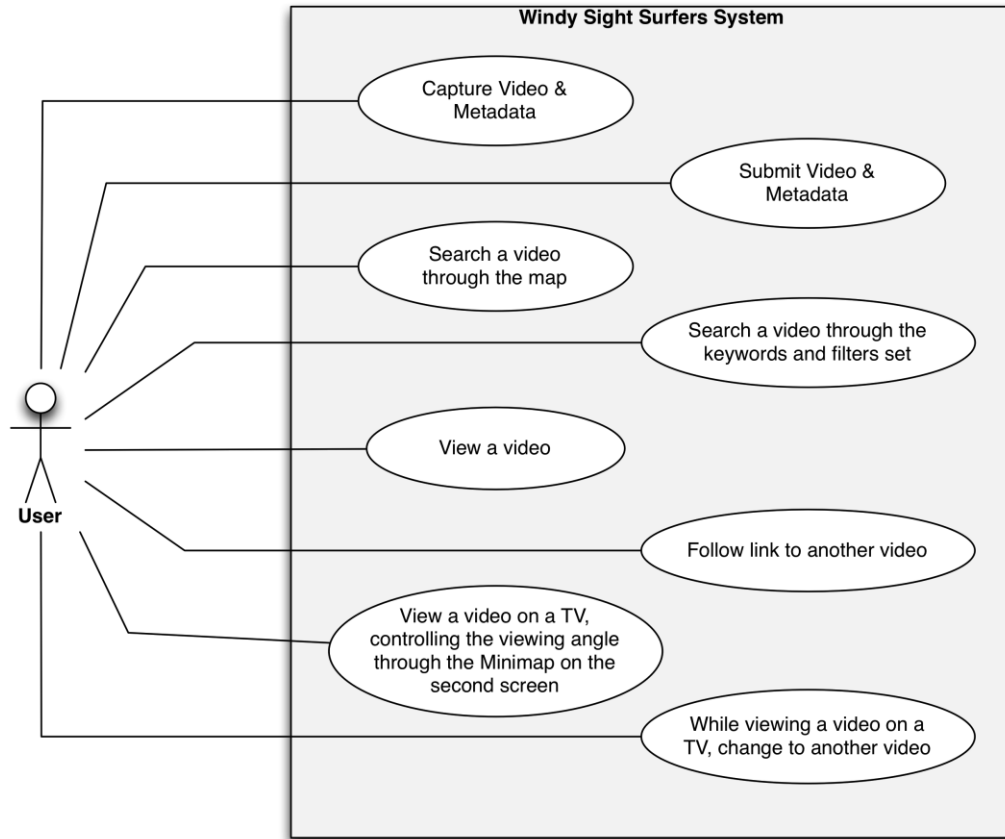


Figure 3.1: Use Cases Diagram

Use Case 1: Capture Video & Metadata

Main Success Scenario:

Through the Capture Mode interface on the mobile application, the user starts the metadata collecting process. Simultaneously, the user also starts the video camera recording. At the end of the recording period, the user simultaneously stops the video camera recording and the metadata recording process.

Use Case 2: Submit Video & Metadata

Main Success Scenario:

Through the Capture Mode interface on the mobile application, the user selects a pre-recorded route and shares it, so it is submitted to the server.

Next, the user selects a video file and submits it to the server, thus completing the submission of the geo-referenced video.

Extensions:

As the video file was not on the mobile device by the time the user submitted the metadata file, the user later accesses the Windy Sight Surfers application (at this time with the video file already in the mobile device) and presses the incomplete submissions notification icon. In the submission menu, the user selects the video file and completes the submission.

Use Case 3: Search a video through the map

Main Success Scenario:

The user starts the search through the map by selecting the search option (e.g. through a magnifying glass icon), and then selecting the “Search Through Map” option. The user then draws paths that resemble the desired routes, or bubbles that represent areas in which the user is interested. Finally, the user submits the query, which completes the search process.

Use Case 4: Search a video through the keywords and filters set

Main Success Scenario:

The user starts the search process of the application. The user then inserts the desired keywords and adjusts the filters to the intended values. Finally, the user submits the query, which completes the search process.

Use Case 5: View a video with the modules designed to enhance the immersiveness on the environment activated

Main Success Scenario:

With the designed modules (e.g. visual, tactile and audio modules) connected to the device the user selects a video in order for it to start reproducing. The modules start to execute coordinated with the video.

Extensions:

- a) With just the audio module integrated, the user selects a video in order for it to start reproducing. The audio component is then reproduced coordinated with the video.
- b) With just the tactile module integrated, the user selects a video in order for it to start reproducing. The tactile component is then reproduced coordinated with the video.

c) With none of the designed modules integrated, the user selects a video in order for it to start reproducing. The bare video is then reproduced.

Use Case 6: Follow link to another video

Main Success Scenario:

During video reproduction, the user detects a hyperlink on the screen indicating an intersection. Selecting the hyperlink results in the change to the intersecting video, starting the new video reproduction at the time associated with the geolocation of the intersection.

Use Case 7: View a video on a TV, controlling the viewing angle through the Minimap on the second screen

Main Success Scenario:

Through the second screen, the user selects a video to start reproduction on the TV, which will result in the interface of the mobile device showing a Minimap of the video being reproduced on the TV screen. By dragging the red rectangle, which highlights the angle of the video being reproduced in the TV screen, or by touching any other part of the Minimap, the angle being viewed on the TV changes.

Use Case 8: While viewing a video on a TV, follow link to another video from the second screen

Main Success Scenario:

When viewing a video through the TV screen, the user detects a hyperlink on the TV screen indicating an intersection. On the mobile device, the user presses the representation of the same hyperlink on the Minimap, which results in the change to the intersecting video, starting the new video reproduction at the time associated with the geolocation of the intersection.

Extensions:

When viewing a video through the TV screen, the user presses the video marker on the map that identifies video trajectories in the mobile device screen and drags it to another trajectory. The video associated with the selected trajectory then starts reproducing at the time associated with the geolocation of the selected point of the route.

3.2 User Registration and Authentication

Users can register in the Windy Sight Surfers system through a standard registration interface, where they provide personal data. Namely, users must provide their birthdate; a username (which must be unique in the system); a password; and an email account. Additionally, for security reasons, users must fill a validation Captcha (url-Captcha), which is a challenge-response test used to determine whether or not the user is human.

When registered, users can login the system by filling a form with their username and password. Additionally, as a means to facilitate the registration and login processes, users can register and login with just a click through their social networks account (such as Facebook and Twitter), since social networks usually provide this functionality (url-FacebookLogin) (url-TwitterLogin).

User Types:

- **Visitors** – are non-registered users. This type of user is able to view and search videos in the system. However, visitors cannot publish videos, or capture their metadata (described in section 3.3 – Video & Metadata Capture). Furthermore, this user cannot profit from advanced features of the system, such as the Emotional Perspective features (described in section 3.8 – The Emotional Perspective). When using the system, visitors are invited to register, in order to become “Registered Users”.
- **Registered users** – besides being able to access the information available to visitors, are able to publish videos, capture their metadata (described in section 3.3 – Video & Metadata Capture), and can profit from advanced features of the system, such as the Emotional Perspective features (described in section 3.8 – The Emotional Perspective).
- **Administrators** – have access to the Windy Sight Surfers DBMS, and are therefore able to manage the information in the system, having read and write permissions over the stored files. The system has not currently an interface dedicated to administrators, mas it is intended to incorporate one in the system, providing easier access and control over the system’s information.

3.3 Video & Metadata Capture

In order to capture 360° videos, a Sony Bloggie Handy Cam1 was used (Figure 3.2). This is a 360° camera available to the general consumer that uses a panoramic lens to capture the 360° of light around it and project it to the video through an upper mirror (Figure 3.3). Through software, this camera allows converting the captured video into a widescreen rectangle image that covers the entire 360° (Figure 3.4).



Figure 3.2: Sony Bloggie Handy Cam1



Figure 3.3: 360° Video, as captured by the Sony Bloggie camera



Figure 3.4: 360° Video, after being converted to a rectangle

Concurrently, the geo-references were captured along the video trajectories with an Android smartphone (creating paths associated to the videos), using the GPS (latitude, longitude, and speed, every half second) and digital compass (orientation angle, whenever direction changes are $> 45^\circ$) sensors. Also, a 3G Internet connection was used to acquire data from the OpenWeatherMap's (url-OpenWeatherMap) weather forecast web service (temperature, weather status, rain, wind speed and orientation). Lastly, the

Android's accelerometer is used to capture and calculate the G-Force values (considering a two-axis system, which fits the desired purposes) during the recording period (every half second). All this data is related to the video being simultaneously recorded by the camera, and hence it is considered the video's metadata. All metadata is stored in an XML file associated with its related video (example of XML metadata file in Annex A).

The reason to not integrate these functionalities (video & metadata capture) in the same device is related to the hardware availability. Currently, most cameras do have neither the sensors nor the software libraries that are common to find in any smartphone. Ideally, this process will be integrated in the same device in a near future.

The Windy Sight Surfers application provides an interface for the capture of the referred metadata and makes use of the Google Maps Android API v2 (url-GoogleMapsAndroidAPI), as well as all the other interfaces of this application that make use of a map system. This interface is accessible from the home screen by clicking on the “Capture Mode” icon (Figure 3.5) and serves three functionalities: collecting metadata, presenting the metadata to the user in real-time (while being captured), and listing and exporting the metadata files.

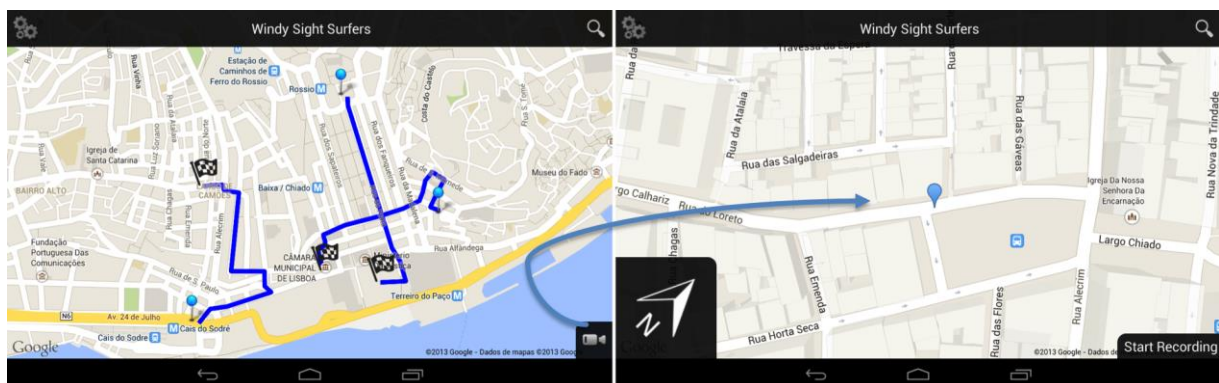


Figure 3.5: Windy Sight Surfers: home screen (left) and Capture Mode (right)

Regarding the metadata collection and presentation, in the Capture Mode interface, a simple and familiar *pannable/zoomable* **map** component fills the background, showing the user his real time position, and serving as a means to present the paths being recorded to the user. On the lower left corner of the screen is a **compass**, which indicates to the user his orientation. As this interface is intended to be as simple and intuitive as possible, apart from these two components, only two more icons compose this interface: a “**Start Recording**” icon, which when pressed starts recording metadata, and an “**Options**” icon, which when pressed opens a pop-up menu with the following functionalities (Figure 3.6):

- **Calibrate the device** – enables the users to assign the “neutral” position of the device during the recording period. For example, if the user is going to have the android device in their pocket while they will be recording the video, this feature enables the device to “understand” its neutral position (Figure 3.6).
- **Go to Current Location** – centers the map on the current location of the device, which is useful when the user is navigating other areas of the map and feels lost.
- **Show Routes** – displays the previously recorded routes that took place on the map area currently being shown on the device (Figure 3.6). By clicking on each of these routes, several options are given to the user, such as edit, share and delete, as it will be explained in finer detail further ahead on this document.
- **List Routes** – is a more conventional alternative to search routes when compared to the “Show Routes” functionality, consisting on a pop-up with the list of routes previously recorded. Each element can be edited, shared or deleted when it is pressed. Apart from the presentation, these three options function in the exact same way as the options given to the user in the “Show Routes” functionality (described further ahead).
- **Clear Screen** – clears all the routes on the screen in order to eliminate clutter when too much information appears on the screen.

By pressing the “Start Recording” icon, and once the device is connected to a GPS satellite, the application starts recording a path. By doing so, the “Start Recording” icon is replaced by a “Stop Recording” icon and a “Minimize” icon appears on the upper right corner, which enables to close the application while keeping running a service on the background (Figure 3.6). The background service enables the user to start recording, close the application, turn off the screen, put the device on his pocket (for example), and concentrate on the video recording without worrying of pressing some key inadvertently. When the user presses the “Minimize” icon, the application closes, and a notification icon appears on the status bar of the Android home screen. As the user intends to stop the metadata recording, it is possible to do so either by re-launching the application, or by tapping the notification on the status bar and pressing the “Stop Recording” icon.

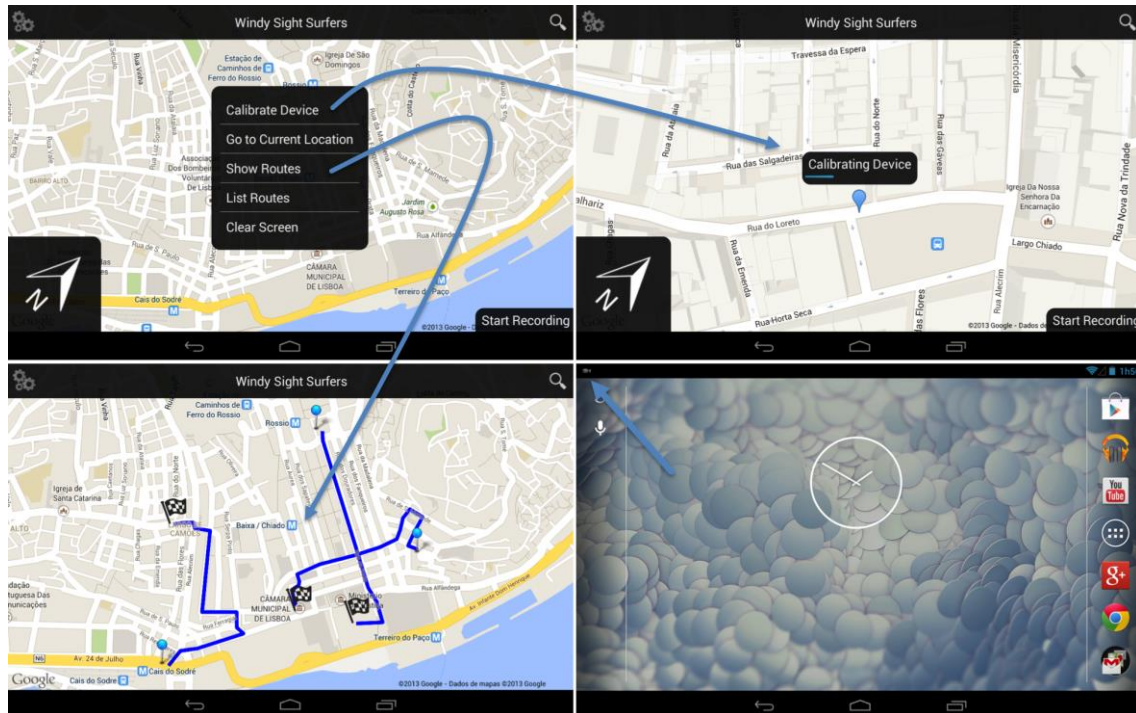


Figure 3.6: Windy Sight Surfers Capture Mode: Options menu (upper left); Calibrating the Device (upper right); Previously recorded routes in the map viewport (bottom left) and Service executing in the background accessible through the upper left corner (bottom right)

Previously recorded paths are represented in the map as a blue line delimited by two markers: a “start marker”, which marks the start point of the path, and a “finish marker”, which marks the end point of the path (Figure 3.6). When the application is recording a path and it is not minimized, the path being recorded is displayed on the map as a red line and it is updated each time the device moves. Also, the finish marker is not a chequered flag, as it is still being recorded, but rather a simple blue marker. In this sense, when the video is being recorded, the path is a line delimited by a “start marker” and a “current position marker”.

When the “Show Routes” option is selected, the map component presents the previously recorded routes contained in the map area being shown on the screen. The routes being showed are updated as the user pans/zooms the map. It is important to note that, on this interface, only the routes recorded on this device are showed. This design decision was taken because this interface is related to the capture and submission of routes. If the user wants to search for all the videos on the server, the main application search features serve that purpose, and are easily accessible from this interface.

3.4 Video & Metadata Sharing

Windy Sight Surfers has its ultimate objective in providing a social platform where users can submit and view each other's routes and experiences. Therefore, the video sharing functionality is embed at the very core of the system, meaning that it is always easy to initiate and prosecute through the video sharing process. The easiest way to initiate the sharing process is by submitting to the server a route that has just been recorded. Therefore, while the user is still in the Capture Mode, it is possible to see the routes recorded on the device by selecting the "Show Routes" option (or the "List Routes" option, which functions in a similar way). As it has previously been described, each route's representation consists of a line delimited by two markers. When a marker of any of these routes is pressed, a pop-up window appears over the marker with "Edit", "Delete" and "Share" options to the user (Figure 3.7).

- The "Edit" option enables the user to override some of the recorded route's automatically generated information. Namely, the weather information given by the OpenWeatherMap web service may be corrected for a more accurate value. Users are given this possibility because the OpenWeatherMap web service collects the weather information from the nearest weather station to the real GPS coordinates. Depending on the circumstances, this station can be far away from the real location, and therefore information might not be as accurate as desired. Also, this option enables the user to add some information to the recorded route. Namely, in the case an Internet connection was not available during the recording period, the user might fill the weather/wind condition.
- The "Delete" option enables the user to delete unwanted routes.
- The "Share" option makes it possible for the user to submit the selected route to the server. When the user selects this option, the first of two steps in video submission is completed, which corresponds to the submission of its metadata file to the user's account. After this step is completed, the second step, which consists in the submission of the video itself, must be completed in order to conclude the submission process. If the video file is already on the device, the user can submit it right away (Figure 3.7). Otherwise, the user might complete his submission later. When the user's account has incomplete submissions, a notification will appear of the upper bar of the application (Figure 3.7). In order to complete the video submission, the user can press the notification icon, which will bring the incomplete submission menu to the front. Finally, the user must assign a name to the video, which will be its main identifier (by default the name will be the video file's name).

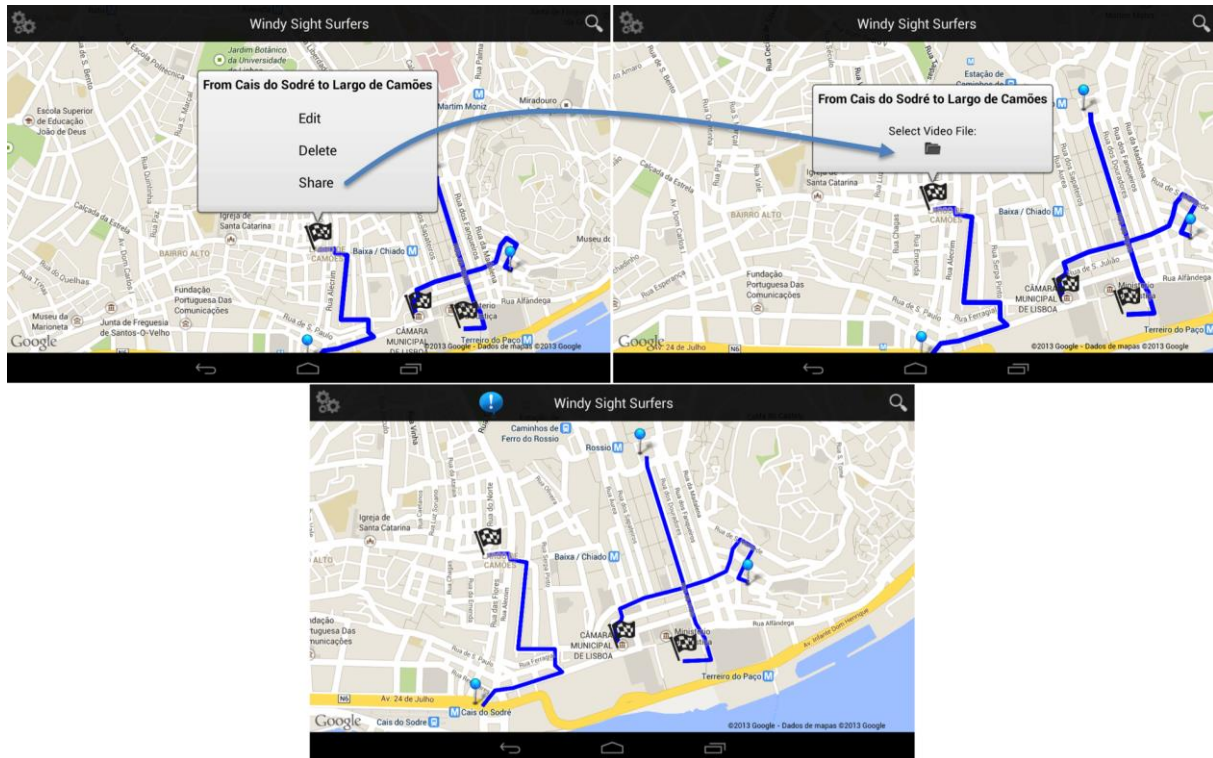


Figure 3.7: Sharing the video’s metadata (upper left) and the video itself (upper right); Notification for incomplete submissions (bottom)

Apart from this sharing method, the options icon, which is present in every menu of the application, features the “Share” option, which transports the user to the “Show Routes” interface, from which the user can carry out the sharing process as described above. This puts the sharing functionality one press away in every use case.

3.5 Video Search

As the searching functionality is at the core of the application, users can start a new search at any point in the application by clicking in the magnifying glass icon and typing the desired search keywords, thus implementing the familiar search approach present in systems such as Google or Youtube. Although it is possible to conduct a search only based on a given set of keywords, there is a set of filters related to the nature of the application (slow/fast, rainy/sunny, day/night, year span, etc.) that allow the user to filter the search results in a more convenient way (Figure 3.8). These filters consist of either specified conditions (e.g. fast, sunny) or captured current situations (current speed). Regarding filters associated with current situations, when activated these conduct a search of the current device situation, accessing the OpenWeatherMap’s web service if the user searches for videos with the current weather conditions, or using the GPS if the user searches for videos associated with the current speed of movement.

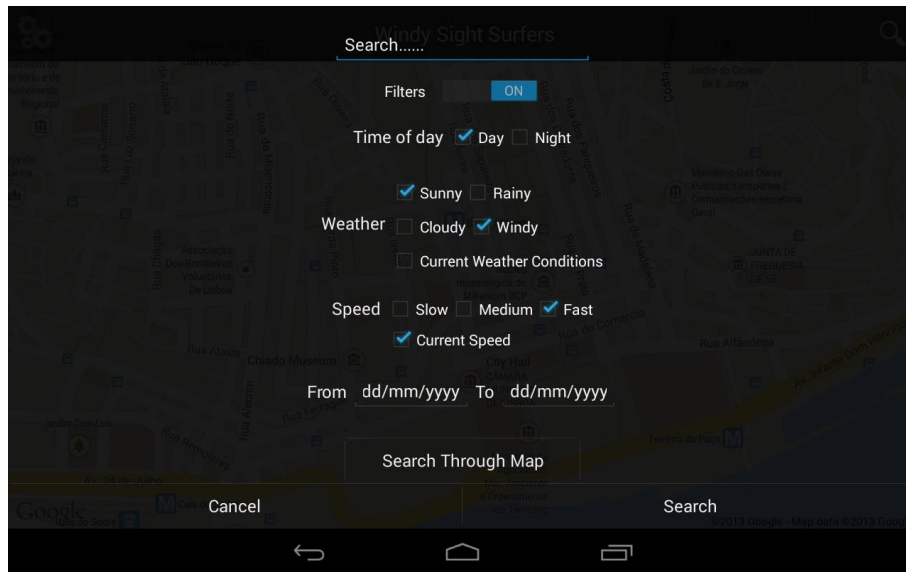


Figure 3.8: Windy Sight Surfers search through keywords and filters

Also, regarding the first Research Question of this thesis (“Do the designed map search features enhance the search process?” – RQ1), in the search window, the user can select the option “Search Through Map”, which allows videos to be searched directly on a map, by using the finger to draw paths that resemble routes in which the user is interested, or drawing bubbles that represent areas in which the user is interested (Figure 3.9).

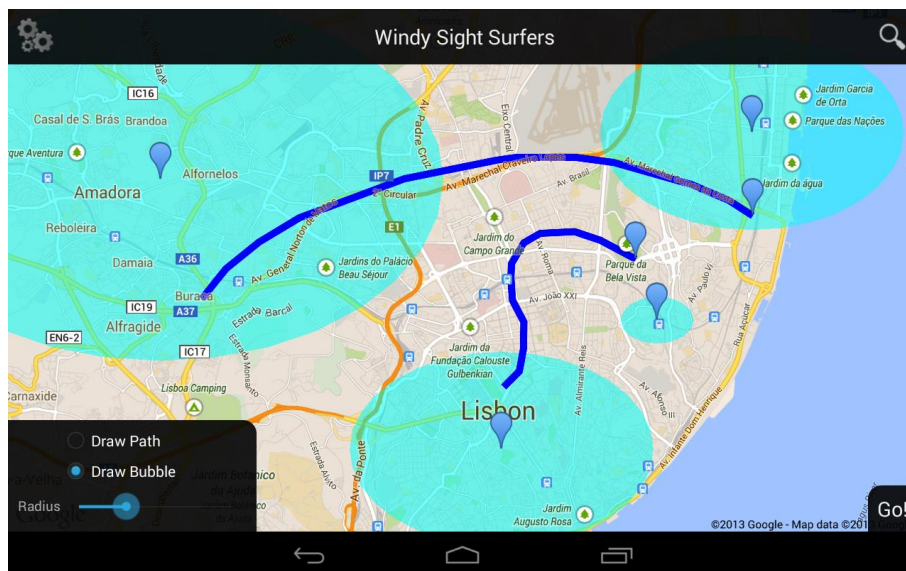


Figure 3.9: Windy Sight Surfers Search Through Map

When the “Draw Path” radio button is selected, the user can draw the desired paths by pressing a finger on the map and moving it according to the user’s desired pattern (while still pressing the finger). When the “Draw Bubble” radio button is the selected one, a slider bar appears under the radio button group, which represents the radius of the

bubbles to be drawn (the higher the slider value, the higher the bubbles radius). In this situation, the user can draw bubbles by pressing the map in the place he desires to be the bubble's center. Each map search element is identified by a marker, which represents the last drawn point of a path or the center of a bubble. Any path or bubble can be deleted, and thus eliminated from the on going search by touching its marker and selecting the "Undo" option in the pop-up window that appears over the marker. When the user drawn all the desired paths/bubbles, he can trigger the search by pressing the "Go!" button.

The process to retrieve the search results depends on which of the methods used to conduct it. If the search was conducted using the keywords/filters system, the result is based on a simple filtering of the videos that comply with the user's demands. This process is similar to the ones used by common search system, such as Google and Youtube. In the case the search was conducted by drawing a bubble on the map, the results are simply based on videos contained on the drawn bubble. Videos are considered contained on a bubble if at least 50% of their path (set of georeferenced coordinates). Lastly, if the user performed a search by directly drawing a route on the map, the search follows an algorithm that divides the path into several bounding boxes and retrieves videos that are also located in the majority of those bounding boxes. Namely, the video must be located in at least 50% of the bounding boxes.

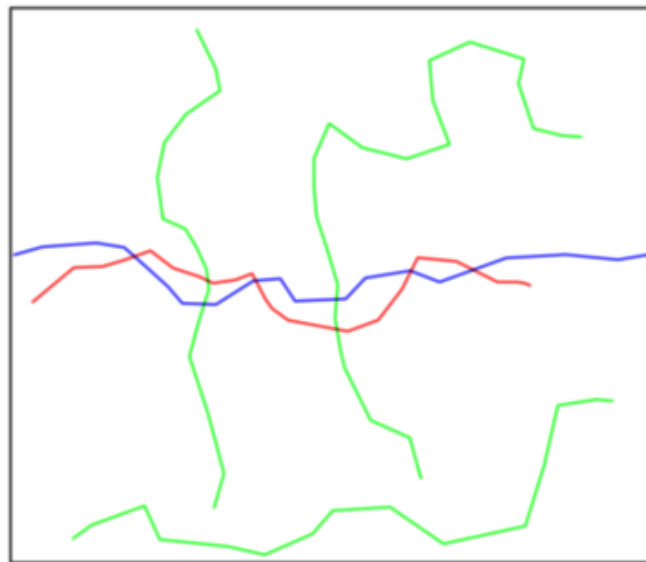


Figure 3.10: Search example, where the blue route was the searched route by the user. If the algorithm used just one bounding box (grey), the search results would include all the green and red routes, when in reality, only the red route represents a relevant search

The reason for this is that if the search was to be related to simple match between any of the coordinates (just one bounding box), the result could be erroneous (e.g. a search of an horizontal path retrieves as result a vertical path perpendicular to the

searched path) (Figure 3.10). As the Search Through Map feature allows conducting searches based on both bubbles and paths at a time, the search system treats each of the drawn objects as individual search objects and, therefore, the final search result in this case consists of the sum of the search results of each of the search objects.

The presentation of the search results can be done in two different ways: the first, is a conventional list of videos, which consists of a cover-flow containing a picture and title of each of the videos (Figure 3.11); the second, and default one, is the presentation of videos in the map as routes, which consist of a line delimited by two markers (as it was previously described in section 3.3 (Video & Metadata Capture). The zoom level is automatically set according to the search results, so that the most results appear on the screen. If the results originate from very distinct geo-coordinates the system will select the most important area of the map to show. As an example, if one hundred results occur in Central Europe and only three results occur in the United States, then the results present a map centered in Central Europe. When dealing with the overload of information when visualizing search results, a clustering technique is applied. In situations where too many results for a small geographic area are returned, the application groups these results in several clusters, which are presented in the map as bubbles (being each bubble's size directly related to its number of elements). As the user zooms in and out the results are automatically grouped/ungrouped. The user can select one of the videos by touching its marker on the map, which pops up an info-box, containing the title of the video and a touchable picture that, when touched, starts the video reproduction (Figure 3.12).

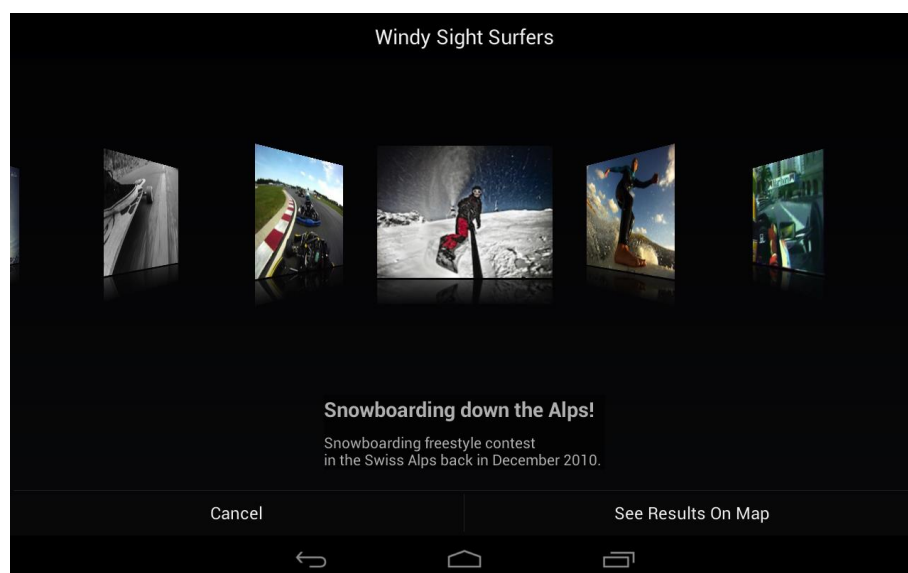


Figure 3.11: Video Search results in Cover-flow presentation

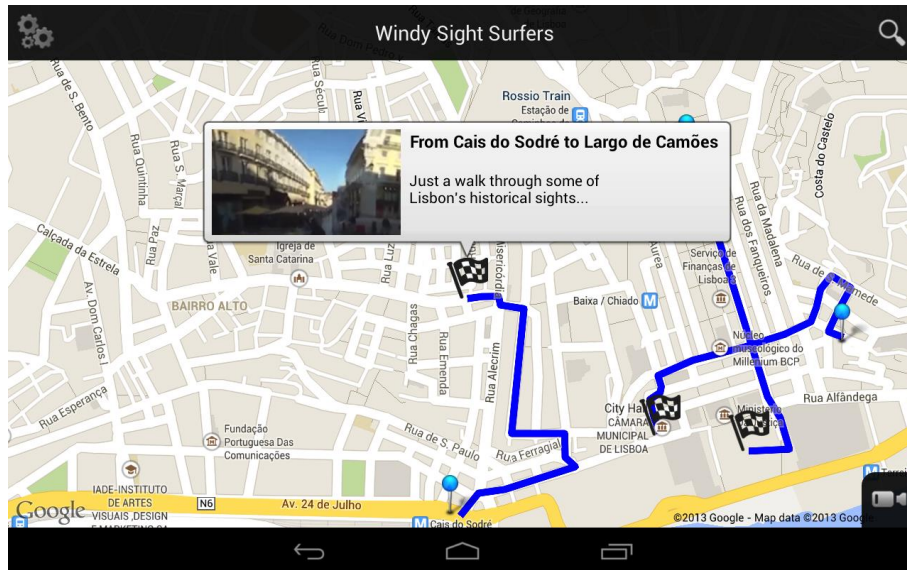


Figure 3.12: Video search results in the map, with info-box highlighting information about one of the videos.

When the application is launched, its initial page displays the videos watched at the moment - “Being Watched Now”. This functionality is part of most applications of this nature, such as YouTube, and usually consists of a plain list of videos. In this case, the interface displays a zoomed out map view presenting a dynamic perspective to the user of the videos being watched at the time, which contains small blue spots on the map, highlighting the map zones that contain videos being watched (Figure 3.13).



Figure 3.13: Videos “Being Watched Now” functionality

These spots are dynamic, which means they are constantly being updated and the user can touch them, which pops up an info-box containing information and hyperlinks to the videos in question. Also, as in every map interface of the application, the map

view is *pannable/zoomable*, which means users can adapt it to present the area in which they are interested. It is important to note that the spots highlight the location of the videos being watched, and not the location of the users watching video. Through the search window, users can always access this page.

In addition to the search methods just described, there is the possibility for the user to search videos through an emotional perspective. Functionalities include the search of videos related to the user's present emotional expression, or his emotional preferences. These functionalities are described thoroughly in the Emotional Module of the application, in section 3.8.

3.6 Perceptual Sensing Features

3.6.1 Visual Sensing in 360° Video

This application targets 360° video. To effectively view and interact with 360° video as an immersive experience, several components are required. Therefore, the 360° videos captured with the Sony Bloggie camera (Figures 3.2, 3.3, and 3.4), are mapped onto a transitional canvas that is in turn rendered around a cylinder, to represent the 360° view and allow the feeling of being at least “partially” surrounded by the video.

When compared to the average aspect ratio of a standard tablet, which for most cases sits somewhere between 8:5 (Google Nexus 7) and 4:3 (Apple iPad), the 360° videos captured by the Sony Bloggie camera have a much wider aspect ratio (128:19). Once the videos are displayed in full screen on the tablet (or any Android device), this means that, if the height of the 360° video is to correspond to the actual height of the device, the video width will be cropped, so there is a need for panning around.

This introduces the second Research Question of this thesis (“Would a full screen pan-around interface increase the sense of immersion ‘inside’ the 360° video?” – RQ2). The approach towards RQ2 considered exploiting Sensor Fusion technology (which is explained in finer detail in section 4.5.1), enabling users to continuously pan around the 360° video in both left and right directions by moving the tablet around, as if the tablet was a window to the 360° video surrounding the user (Figure 3.14).

Although the option to pan the video by moving the device can be a very realistic and immersive approach to pan around 360° videos, there might be some situations where the user is not willing to move the device, such as when the user is seated on a couch. In order to suit both scenarios, users can pan around the video without having to move, by making the entire screen consists of a drag interface, as in the previous approach (Noronha et al, 2012). By swiping to the left or right with one finger over the video view, the video angled is panned accordingly (Figure 3.15).



Figure 3.14: Pan around the 360° video in both left and right directions by moving the tablet around



Figure 3.15: Drag interface

To increase immersion during video playback, the media controller interface is hidden during video reproduction and appears only when the user taps the screen (short press). This interface appears temporarily on the bottom of the screen and contains the standard media controls for the video being reproduced (play, pause, fast-forward, rewind).

3.6.2 Tactile Sensing Through a Wind Accessory

The Wind Accessory was the first of the developed Perceptual Sensing Features (Figures 3.14 and 3.15). An Arduino (url-Arduino) based prototype was designed and built, and is intended to be mounted on the back of the tablet.

Considering the third Research Question of this thesis, defined as “Does wind contribute to increasing realism of sensing speed and direction in video viewing?” (RQ3), the purpose of this device is to blow wind to the viewer during video

reproduction, creating a more realistic perception of speed and movement, and thus increasing the sense of presence and immersion. Therefore, the Wind Accessory operates its fans in real time according to messages received from the Windy Sight Surfers application. When a video is to be reproduced, a new Wind Accessory communication session is initiated, and the application sends messages to the Wind Accessory specifying the frequency the fans are to be rotated (one message per second). These values are calculated according to information contained in the video's metadata file. More specifically, the wind values take into account the wind speed and orientation during the video's recording, the speed the user was travelling during the recording, and the angle of the video being viewed during video reproduction. The ways these factors are taken into account to calculate the values that will be sent to the Wind Accessory involves a three-step normalization process, which is described next:

The first step is the normalization of the speed values contained in the video's metadata file according to the wind speed registered by the OpenWeatherMap web service (in MPS – Meters per Second). Taking into account the MPS wind speed values scale, and the respective “Effects on Land” scale, which are comprised in the Beaufort Scale (Table 3.1), a pairing was established between the wind speed values, and a factor (0.8, 0.9 or 1), by which the wind values are multiplied. More specifically, if the wind speed value is less than 8, than the wind factor is 0.8; if the wind speed value is between 8 and 17, than the wind factor is 0.9; if the wind speed value is greater than 17, than the wind factor is 1. Before the video starts playing all the values in the video's metadata file are normalised by this factor value. This step is done in order to take into account the wind speed during the video's recording.

An example of benefit of this step is the case where two similar videos of the exact same path are recorded at the exact same speed in different occasions. In the first, it was a sunny day with very low wind speed values, whereas in the second video it was a very windy day, and thus with high wind speed values. In this situation, this filter enables the user to notice that one of the videos was recorded in a windy environment.

The second step normalizes the values obtained in the first step. This step is needed because the way the fans are operated in the Arduino platform is through Pulse Width Modulation (PWM), and the values PWM receive must be between 0-255 (described thoroughly in section 4.5.2). Therefore, before the video starts playing all the values are analysed and normalized to fit in the 0-255 range. This is done prior to the video reproduction as a means to save resources, as the application can be hardware demanding during video reproduction.

Beaufort Number	Description	MPS	Effects on Land
0	Clam	0 - 0.2	Calm. Smoke rises vertically.
1	Light air	0.3-1.5	Smoke drift indicates wind direction. Leaves and wind vanes are stationary.
2	Light breeze	1.6-3.3	Wind felt on exposed skin. Leaves rustle. Wind vanes begin to move.
3	Gentle breeze	3.4-5.4	Leaves and small twigs constantly moving, light flags extended.
4	Moderate breeze	5.5-7.9	Dust and loose paper raised. Small branches begin to move.
5	Fresh breeze	8.0-10.7	Branches of a moderate size move. Small trees in leaf begin to sway.
6	Strong breeze	10.8-13.8	Large branches in motion. Whistling heard in overhead wires. Umbrella use becomes difficult. Empty plastic bins tip over.
7	High wind, moderate gale, near gale	13.9-17.1	Whole trees in motion. Effort needed to walk against the wind.
8	Gale, fresh gale	17.2-20.7	Some twigs broken from trees. Cars veer on road. Progress on foot is seriously impeded.
9	Strong gale	20.8-24.4	Some branches break off trees, and some small trees blow over. Construction/temporary signs and barricades blow over.
10	Storm, whole gale	24.5-28.4	Trees are broken off or uprooted, structural damage likely.
11	Violent storm	28.5-32.6	Widespread vegetation and structural damage likely.
12	Hurricane force	>32.7	Severe widespread damage to vegetation and structures. Debris and unsecured objects are hurled about.

Table 3.1: Beaufort Scale – Correspondence between MPS wind values and Effects on Land

The third, and last, step occurs during video reproduction and relates to the angle the user is viewing at each moment while viewing the video. If in the real (recording) situation the person is moving against the wind direction, the user will feel much more wind resistance when compared to the situation where the person is moving along the wind direction. This situation also happens when the person moving turns their head around: as the human hears' shape allows sound coming from the front to be much more audible than sounds coming from the back, when a person turns his head against the wind direction, the hearing perception is that the wind is much higher than when the head is turned to the wind direction. In order to mimic this characteristic, with the intent of making the experience more immersive, before each message is sent to the Wind Accessory, the value obtained in the second step that is about to be sent goes through a last normalization. Before the message is sent, the angle of the video being video is

taking into account so that if the user is viewing the angle of the video that corresponds to the wind orientation during recording, then the value is multiplied by 1; if the user is viewing the angle of the video that is the opposite to the wind orientation during recording, then the value is multiplied by 0.6; if the user is viewing an angle of the video that is approximately between 90° of the wind orientation during recording, then the value is multiplied by 0.8.

After this three-step normalization process, during video reproduction each value is sent to the Wind Accessory, thus creating a wind perception of the video being viewed. When the video reproduction ends, the Wind Accessory communication session is also ended, which stops the Wind Accessory's fans. The Wind Accessory's hardware architecture is described thoroughly in section 4.5.2.

3.6.3 Auditory Sensing: Spatial Audio

As it was described in section 3.6.1 (Visual Sensing in 360° Video), videos are reproduced as if the tablet was a window to the 360° video surrounding the user, meaning that a specific video angle is being viewed at a time. In contrast, when a video is recorded the sound is recorded accordingly to the orientation of the camera, which means that, if not addressed, the sound component of the video is reproduced statically (the sound is always the same regardless of the angle being viewed). This introduces the fourth Research Question of this thesis: "Does a 3D mapping of the video sound allow for easier identification of the video orientation while it is being reproduced?" (RQ4).

Therefore, aiming to allow the user to experience the video sound in the same orientation as when it was captured (through sound) while it is being reproduced, the video's sound is mapped onto a 3D Sound Space. In this sound space, the sound source's position is associated to the front angle of the video (Figure 3.16) and, therefore changes accordingly to the angle of the video being visualized. That is, if the user is visualizing the front angle of the video, the sound source will be located in front of the user's head; if the user is visualizing the back angle of the video, the sound source will be located in the back of the user's head. As videos have 360°, the sound source's location changes over a virtual circle around the user's head (Figure 3.17). These sound source's location changes are only relative to the horizontal space. Although the space is a 3D sound space, the vertical position was never changed, and was always kept at the human hear level. This decision was based on the fact that the captured 360° videos also only enable horizontal panning.



Figure 3.16: 3D Audio: Red rectangle representing the video viewing viewport (movable throughout the 360°). Audio source location associated with the front video angle (fixed association).



Figure 3.17: 3D Audio: source location changing around the 360° video viewing. Grey stripe on top represents video trajectory direction

As it will be described in section 5 (User Evaluation), during the user evaluation users provided feedback on the best values for the virtual “distance” between users’ head and the sound sources. Namely, user preferred to dispose the sound source between one and three “virtual” meters, which means the radius of the virtual circle must be within these values.

3.6.4 Auditory Sensing: Cyclic Doppler Effect

Considering the fifth Research Question of this thesis, defined as “Can a controlled use of the Doppler Effect increase the movement sensation while viewing videos?” (RQ5), a sound experiment was conducted to study this condition.

The Doppler Effect can be described as the change in the observed frequency of a wave, occurring when the source and/or observer are in motion relatively to each other. As an example, this effect is commonly heard when a vehicle sounding a siren approaches, passes, and recedes from an observer. Given the fact that people inherently associate this effect to the notion of movement, an experiment was carried out to see if a controlled use of the Doppler Effect could increase the movement sensation of users while viewing videos. In order to do so, a second sound layer was added to the video, which cyclically reproduces the Doppler Effect in a controlled manner. Similarly to the Spatial Audio feature, the sound corresponding to the Doppler Effect Layer is also mapped onto a 3D sound space. In the basis of this sound layer is a sound that is

reproduced cyclically and approaches, passes, and recedes the users' head (from the front to the back) (Figure 3.18).

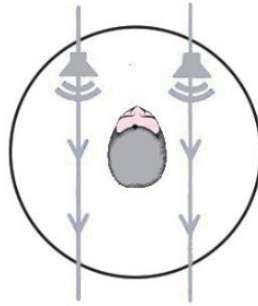


Figure 3.18: Doppler Effect: Audio changes cyclically as in grey paths

Regarding the Doppler Effect, there are several aspects that influence the intensity of the movement sensation. Especially, the intensity of the movement sensation is affected by the intensity (volume) of the sound, and the rate at which it is reproduced. Also, the sound itself used to reproduce the effect can be of great importance, as some sounds might be more effective (create a stronger movement sensation), but also more intrusive (interfere with the main sound layer). With respect to the **rate** at which the sound is played, this value is set while playing and it varies during playback, according to the speed values stored while capturing the video (the value is updated every three seconds). In other words, the higher the speed, the higher the intensity of the Doppler Effect. Concerning the **sound** used to reproduce the Doppler Effect, as it will be described in section 5 (User Evaluation), several experiments were conducted with the intent to find out the right parameters. Several types of sounds were experimented, aiming to find the sounds that create a stronger movement sensation, while not being intrusive. This proven to be complex problem, as some of the most effective sounds were also considered the most intrusive, which resulted in a trade-off situation that is not easily resolved and might be grounds for further research. Also, the threshold level of the volume of the Doppler Effect sound layer's sound sources was measured, as well as the optimal sound value.

One of the side effects of this approach might be to try to alert the user for movement when there is little/no movement. This can dramatically change the effectiveness of this feature by turning it into something obtrusive rather than beneficial. Therefore, this problem was also analysed with the intent to find if there was a minimum amount of movement required for the Doppler Effect to become beneficial. As results shown (section 5), there is a minimum amount of movement required, which led to the development of a high-pass filter that added the requirement for a minimum amount of movement in order for the Doppler Effect simulation to execute.

3.7 Context Awareness Features

Striving to increase immersion through context awareness, and considering the sixth Research Question of this thesis, defined as “Do the designed context awareness features contribute to a more immersive environment?” (RQ6), when the video is being reproduced, an overlay is always present above it, which enriches the video with additional features that provide information from the video’s metadata (Figure 3.19). These features are all user selectable (can be activated/deactivated through the options menu) and are divided in two categories:

- 1) Permanent Information Features, which relate to items of information that are permanently present on the screen throughout the video reproduction;
- 2) Momentary Information Features, which relate to items of information that appear momentarily on the screen and are related to specific portions of the video.



Figure 3.19: Video Context Awareness: Video being reproduced with the overlay, displaying the orientation (top right), speed (bottom right) and G-Force (bottom left) values

3.7.1 Permanent Information Features:

The **Video View Area** (Pizza-slice) is a circular “pie-chart”-like interface, which uses a red angle to highlight where the user is looking at (Figure 3.19, top right corner). Furthermore, the pie-chart interface contains a north indicator and a video center indicator, which are synchronized with the map movement (known through the geo-information in the metadata file) and change whenever the angle of direction changes, allowing users to keep their sense of direction. Three variations of this functionality were designed and are user selectable: 1) video center up (fixed on top) – for alignment with the author’s perspective while filming (Figure 3.20a); 2) north up – to align with

the geographic north (Figure 3.20b); 3) video current view, or orientation, up – to align with the video perspective being shown (Figure 3.20c) (variation similar to the one adopted by Google Street View). Furthermore, the Video View Area interface also serves the purpose of quick video angle navigation, by dragging the red area to the desired position (Noronha et al, 2012).

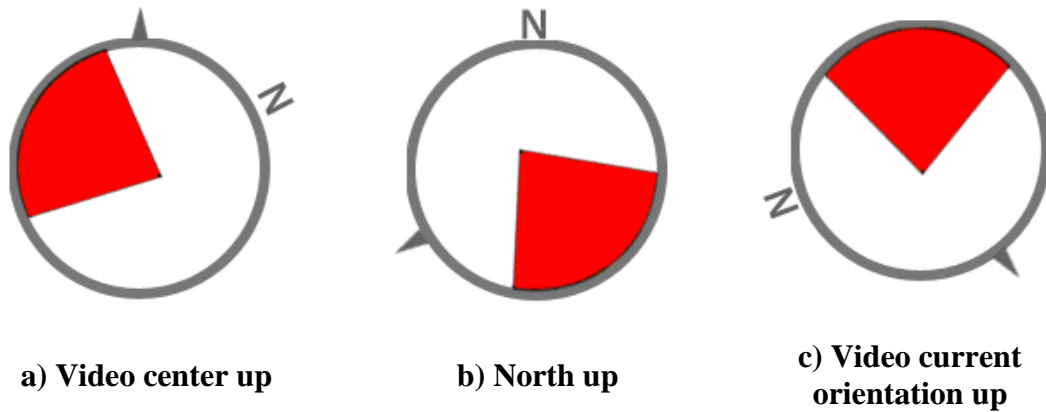


Figure 3.20: Different compass modes

On the bottom right corner of the screen is a **Speedometer** (Figure 3.19), which presents the (GPS captured) speed information contained in the video's metadata file. An early version consisted of a computer simulation of an analog speedometer, but was considered too intrusive by early test results, which led to the design of a simpler digital speedometer.

On the bottom left corner of the screen is a **G-Force meter** (Figure 3.19) that presents the G-Force withstood by the filmmaker while filming. The G-Force information is calculated by the metadata capture functionality and stored in the video's metadata file. The way this information is presented to the user is through an interface commonly used in sports videos, where the G-Force value is represented by a red dot, which appears in the resulting position of a two dimensional axis graph, being that the vertical axis represents G-Force values ranging from the user towards the device and the horizontal axis represents the horizontal range of G-Force values (from left to right). Below the graph there is a value, which represents the overall G-Force value.

3.7.2 Momentary Information Features:

The **Video Landmarks** consist of visual and vibratory notices to notify the user about relevant happenings on specific moments, such as the maximum G-Force value in the video. During the video search process, some information is retained regarding the user's choices, filters and keywords, and while the video is being reproduced this data is taken into account. So, when the user searches for 'fast' or 'speedy' videos, a maximum speed notice is presented during video viewing, or when the user searches for

‘High G-Forces’ videos, a maximum G-Force notice is presented during video viewing. Alternatively, through the preferences menu, the user can specify which of the momentary information items he is interested in, by selecting the momentary information items to be presented from a checklist.

The Windy Sight Surfers application supports hypervideo, by providing spatiotemporal hyperlinks, which can be defined in any angle/position and at any time of the 360° video. These hyperlinks are touchable by the user and are represented by coloured rectangles, which follow the video’s angle rotation, meaning they are associated with a specific angular position of the video (Figure 3.21).



Figure 3.21: Hyperlinks in 360° video (only context awareness feature activated in this picture): POI (left) and Movie (right)

Currently, the following kinds of hyperlinks are implemented:

- **POI's (Point of Interest)**, which mark points of interest of any kind to the user, such as an underground station or a restaurant. Most importantly, this kind of hyperlinks includes a link to a related webpage, where users can access additional information regarding the POI.
- **Crossing Trajectories**, which point out zones where there are crossing trajectories, allowing the user to navigate to the videos associated with the crossing trajectories by just pressing the hyperlink, and thus not having to leave the video and select the crossing trajectory through the map (which allows the user to switch between videos and trajectories in a more immersive way). When considering a highly populated video system, this feature can be of great value, for example, it might allow navigating through cities using 360° videos.
- **Movies**, which allow certain parts/places of the video to be associated with certain movie scenes. When this hyperlinks are touched, the video pauses and a pop-up window appears, reproducing a scene of a movie that was recorded on the same place as the current video. This allows the users to have an historical perspective on cinema related to the places he is viewing on 360° videos.

Contrary to conventional video, in 360° video most of the content is outside the viewport. Therefore, in addition to have hyperlinks over the video, some techniques must be designed and implemented in order to provide *Link Awareness* to the user. This is vital to this kind of video, because without link awareness users might not even notice hyperlinks for the most of times. In this application, Link Awareness is provided to the user through the following functionalities:

- **Hotspots**, which consist of labelled coloured balloons on the video, providing additional info about the video object and link.
- **Hotspot Indicators**, which consist of marks that appear on the Lateral edges of the window to alert the user (provide awareness) about links that are outside the viewport. When a hotspot indicator appears, its lateral position indicates the hyperlink's vertical position, and the size of the hotspot indicator indicates the distance to the hyperlink (the closer the hotspot is to the current viewport, the bigger the hotspot indicator).

This overlay comprises a set of features that intend to increase immersion through context awareness, when the video is being reproduced.

3.8 The Emotional Perspective

With the aim of providing a more engaging experience to the user, an emotion recognizer system was incorporated into the application. The Emotion Recognizer is based on a Facial Expression Recognition Framework (Mourão et al. 2013) that is used in the context of the Windy Sight Surfers application as a means to recognize the user's emotions when visualizing video and using the application. The referred framework recognizes eight emotions (Neutral, Anger, Contempt, Disgust, Fear, Happiness, Sadness, Surprise), which include the six basic emotions that were recognized across cultures in one of the major studies on emotional expressions performed by Paul Ekman (Ekman, 1992).

As this work fits on the mobile environment context, the front-facing camera present in most mobile devices can be used to capture the user's facial expression while viewing videos. Once captured, the Emotion Recognizer processes the captured images and an expression is associated with each one of them. These results are stored in a database, which is later used in several features of the application with the aim of improving the user experience.

The Emotion Values database includes information related to the videos, such as the emotions more frequently associated with each video; and information related to the video – user relation, such as the emotions exteriorized by the user while viewing each

video. It also includes user information, such as the most prevalent emotions associated with that user (regarding all the videos she has watched). Since users need to log in the application when starting interacting with it, the emotions recognized by the system can be easily associated with each user's account.

At present, the emotional information has three applications in the created environment: this information is used in features related to a Video Emotional Cataloguing and Access (on the map and through search); in features that may influence the control flow of the environment; and it was used in the user evaluation of the environment itself, based on the emotional impact (an emotional evaluation module was developed).

In this context, and considering the current state of development of this module, the seventh Research Question of this thesis was defined as follows: "Do users consider the emotional perspective relevant in the access and search of videos?" (RQ7).

The following section presents a detailed description of these three applications of the Emotion Recognizer.

3.8.1 Video Emotional Cataloguing

3.8.1.1 EmoMap

As a mean to improve the search capabilities of the application, the EmoMap feature was designed. This feature is associated with the "search through the map" functionality (section 3.5), which enables users to search for videos through the selection of areas and paths on a map. When the EmoMap feature is activated, a semi-transparent overlay is added to the map, with the intent of colouring it according to the dominant emotions of the videos in each area, and thus enabling the user to identify the map zones that contain videos associated with specific emotions (Figure 3.22).

In this approach a representation of emotions based on colours was adopted, like the Plutchik's wheel of emotions (Figure 2.14b) (Plutchik, 1980).

Also, the user can search for videos associated with specific emotions. The application includes two methods to search videos with emotion filtering. The first, and more conventional one, enables filtering through the selection of the desired emotions from a list of checkboxes (Figure 3.23). The second method requests the user to express an emotion in front of the camera of the device through the "Express Emotion" option, analyses that emotion and retrieves videos associated with it to the user.



Figure 3.22: EmoMap

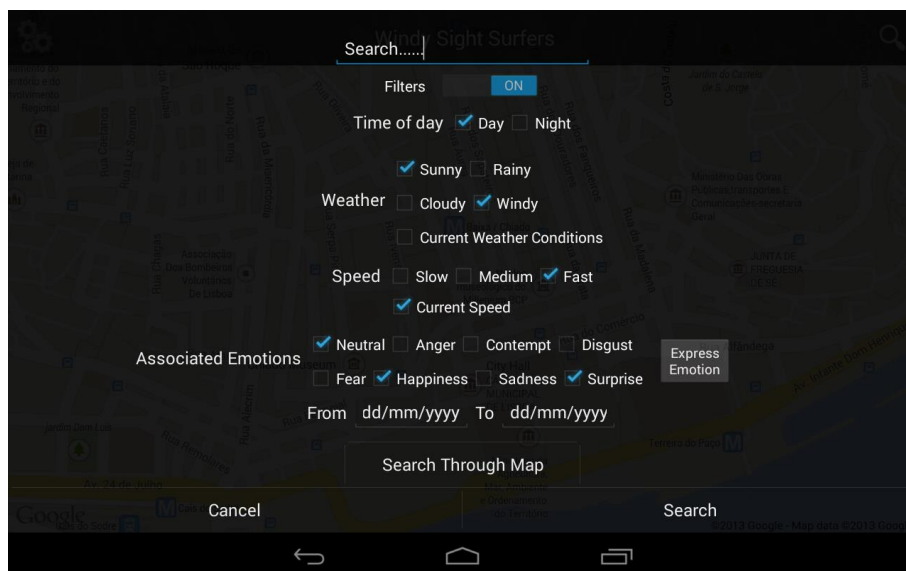


Figure 3.23: Windy Sight Surfers search through keywords and filters

The presentation of the search results can be done in two different ways: the first and default one, is the presentation of the videos in the map, in which the map zones that contain videos predominately associated to the searched emotion are highlighted in semi-transparent blue bubbles, similar to the presentation of search results highlighted in Figure 3.12. The user can select one of the videos by touching its marker on the map, which pops up an info-box, containing a picture and the title of the video; the second option for the results presentation, is a more conventional list of videos, which consists in a cover-flow containing a picture and title of each of the videos, similar to the presentation of search results highlighted in Figure 3.11.

3.8.1.2 EmoMe

EmoMe is an interface that presents the user statistical information regarding his personal emotional history in the application. In addition to the information related to the previously viewed videos, several charts and graphs present video trends, as well as recommendations. In order to be able to analyse the real effect of this feature, the application is required to rely on a highly populated 360° video database, which is yet to be achieved. This means that the design of this feature could not be tested thoroughly at the time of this writing. Therefore, although the feature is already implemented, it needs further testing (when a highly populated 360° video database is available) in order to be able to learn about its perceived usefulness, satisfaction and ease of use, which may eventually led to the refinement of its design and further developments.

The recommendation system takes into account, not only the user data, but it crosses user's information by also taking into account data from "similar" users (with similar preferences/emotional history). These "user preferences" are not specified by the user, but rather inferred by the system as the user views videos. As the number of videos viewed by users increases, the preferences system is expected to become more precise; being the system's rational as follows: as a user views a video, its emotional information is stored and associated with that specific user and video. As more users view the same video, the system can calculate the most prevalent emotions associated with that video, relating not only to each specific user, but also to the entire community. Also, as the number of videos a user views increases, the system can infer which are the user's video emotional preferences by analysing the user's history. Using this information, the system chooses other videos that fall in this category (user preferences), and present the user with recommendations.

3.8.2 Emotion Driven Control Flow

Based on the user's emotional profile, as the user is viewing a video and the Emotion Recognizer is recognizing user emotions, suggestions can be presented to the user if, e.g. the emotion recognizer detects a negative pattern regarding user emotions. These suggestions will use the same recommendation system described in section 3.8.1.2 (EmoMe), and are presented as hyperlinks on the video, which upon being touched stop the video reproduction and open a cover-flow with a list of video recommendations, similar to the search results list described in section 3.5 (Video Search). This feature intends to improve user satisfaction by minimizing the time users spend viewing undesired video content.

3.8.3 User Evaluation Based on Emotional Impact

As a mean to conduct a more representative evaluation of the environment, the Emotion Recognizer is used as an evaluation tool. This means that the system was used itself as a tool to collect data during the entire user evaluation process. In each video/menu the dominant emotions are logged, which allows analysing the user's dominant mood during each stage of the evaluation, as well as during the entire user evaluation process. The results of the use of this feature are presented in section 5.2.7 (Evaluating the Emotional Impact of Windy Sight Surfers).

3.9 Interaction with TVs & Wider Screens using Second Screens

Considering the context of this work, and striving to increase the immersiveness of the environment, the application has the capability to interact with wider screens, such as TVs, and take advantage of their screen size, thus introducing the eighth Research Question of this thesis: “Does the interaction with TVs & wider screens, with video in full screen and additional content and navigation control in a second screen, contribute to a more immersive environment?” (RQ8).

In fact, for the context of this thesis, an ordinary computer was used as the host of the “Future Interactive TV” application, but will be referred to as TV throughout this section. Building on previous work (Noronha, 2012; Álvares, 2012), it consists of an HTML5 web application running in a standard web browser supporting WebGL (url-WebGL), which allows viewing 360° video integrated in a map system through a widescreen.

At the present scenario, the computer acts as a TV, so no mouse or keyboard interaction is used to control the web application (which is designed to be used in full screen). All the interaction with the computer is achieved via the Windy Sight Surfers mobile application. As the system strives to be as seamless to the user as possible, the connection between the mobile device and the TV has as its only requirements that the application is installed in the mobile device; the computer (acting as a TV) has a web-browser with the named web application running; and there is a wireless network to which both devices are connected.

The connection process is initiated by the system. Whenever the mobile application detects a TV in a short range the user is given the possibility to extend the application to the TV screen. If the user accepts the connection, the mobile application rearranges itself so that the videos are reproduced in the TV screen, and the mobile device's screen is used for second-screen purposes. In this situation, as the TV becomes responsible for reproducing the videos, the mobile application has two main roles in the environment:

control the TV, and displaying additional metadata, related videos, etc. Next, the application's second screen functionalities are presented.

3.9.1 Minimap

When a 360° video is being reproduced in the TV, the mobile application shows a minimap, which consists in a projection of the full 360° of the video that provide increased perception and control of the whole video image (Figure 3.24). As an overlay on the minimap, a red rectangle highlights the angle of the video being reproduced in the TV screen, and this angle can be changed by dragging the rectangle or by touching any other part of the minimap. Therefore, in the second screen context, the minimap drag interface is the main tool to control the viewing angle (pan around the 360° video). An additional method to pan around the 360° video was implemented, and consists in turning the mobile device as if it was a steering wheel. When this method is activated, users can change the view angle by holding the mobile device and turning it to the left and right, which will change the angle of the video being displayed.

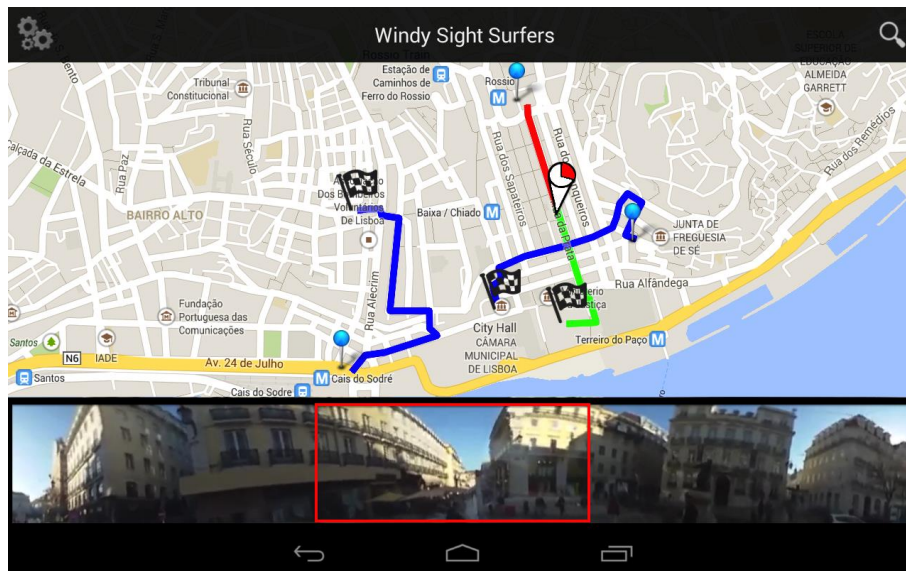


Figure 3.24: Windy Sight Surfers Second Screen application

Similarly to the stand-alone application, the media controller interface is hidden during video reproduction and appears when the user taps the minimap (short press). This interface appears temporarily on the bottom of the screen and contains the standard media controls for the video being reproduced (play, pause, fast-forward, rewind).

3.9.2 Hyperlinks

The approach for presenting hyperlinks in the TV screen is the same used to present them in the standalone mobile application (Hotspots and Hotspot Indicators, described in section 3.7 (Context Awareness Features)). However, in order to make it easier for the user to navigate through the hyperlinks shown in the video on the TV

screen, they also appear in the mobile application's minimap. This feature serves another purpose: once the minimap displays the 360° of the video at any time, the user has easy access to hyperlinks that are out of sight on the angle being viewed on the TV screen. As the mobile application is used in this context as a second screen, it is expected that the user is focusing on the TV screen for the most of the time. Therefore, as a means to increase the user awareness of those hyperlinks, whenever a hyperlink first appears outside the TV viewport, the mobile device emits a short vibrating alert. To select a hyperlink the user must long press its hotspot, which will activate the hyperlink's content. The hyperlink's selection must be based on a long press because a simple press on the screen could possibly create false positives, for example when the user is panning around the video through the minimap.

3.9.3 Geographical Navigation & Orientation in the 360° Videos and Maps

While videos are being viewed on the TV, the mobile application has the ability to present a map, which identifies video trajectories (Figure 3.23). The path corresponding to the current video is shown in green (colour that humans are more sensible to, and associated with growth, life and action), while other trajectories in the area are shown in blue (colour that humans are less sensible to, and associated with calmness). When a trajectory has been viewed it is painted red, to help users track which ones were already viewed. On the map, a marker signals the geographical position of the video at any given time. Also, instead of using the Google Maps standard marker, a dynamic marker was designed, which contains an interface similar to the Video View Area feature, described in section 3.7 (Context Awareness Features) that indicates which area of the map corresponds to the angle of the video being viewed at each moment (Figure 3.24). As the video progresses in time, this marker advances through the video's route, and when the user changes the video's viewing angle, the marker's Video View Area is also updated. By touching or dragging the marker over trajectories, the user can navigate to the corresponding video and time.

3.9.4 Related Videos

While videos are being viewed in the TV, the user can also use an interface in the mobile application that shows a touchable list of related videos. This list consists not only of videos that relate to the video being viewed, but it also contains videos that match the user preferences. The priority levels are as follows: the first videos of the list are those which relate to both the user preferences and the video being viewed; next in the list are the ones related to the video being viewed; lastly, the videos related only to the users' preferences are suggested.

Chapter 4

System Implementation

Having the design and functionalities of Windy Sight Surfers been presented in Chapter 3, this chapter focuses on the implementation aspects of it. Nowadays, mobile platforms, such as Android, offer a wide range of tools that allow the computing community to explore and enhance these platforms. Also, Web technologies are incredibly vast, and ever improving. However, due to the exploratory nature of this thesis, several challenges emerged during implementation. As such, they are analysed in this chapter. These challenges relate to both the functionalities design (conceptual challenges), and to their implementation (implementation challenges).

As it will be explained throughout this chapter, components such as the Wind Accessory proved to be big implementation challenges, because, although some of these platforms (such as Android or Arduino) can be considered “mature” at the date of this writing, integration platforms (such as the Android Open Accessory Protocol) are still at an embryonic stage. In addition, the functionalities related to the Auditory Sensing (3.2.4.3 and 3.2.4.4), Emotional Perspective (3.2.6), and Interaction with TV’s & Wider Screens (3.2.7) components required the use of several Web technologies that are themselves experimental, which created difficulties.

The first two subsections focus on the System Architecture and the System Development Methodology. Once these topics have been described, the major decisions that were taken during the design and implementation processes, regarding the major features of the Windy Sight Surfers application are discussed between sections 4.3 and 4.8. For each of these features, design options are analysed, used technologies are explained in detail and algorithms are thoroughly examined.

4.1 System Architecture

Windy Sight Surfers is based on a web service architecture where frontend applications request the data from a backend (Figure 4.1). The backend contains a relational database, which stores all the information related to the system, and provides

several web services that allow frontend applications to access the information they need.

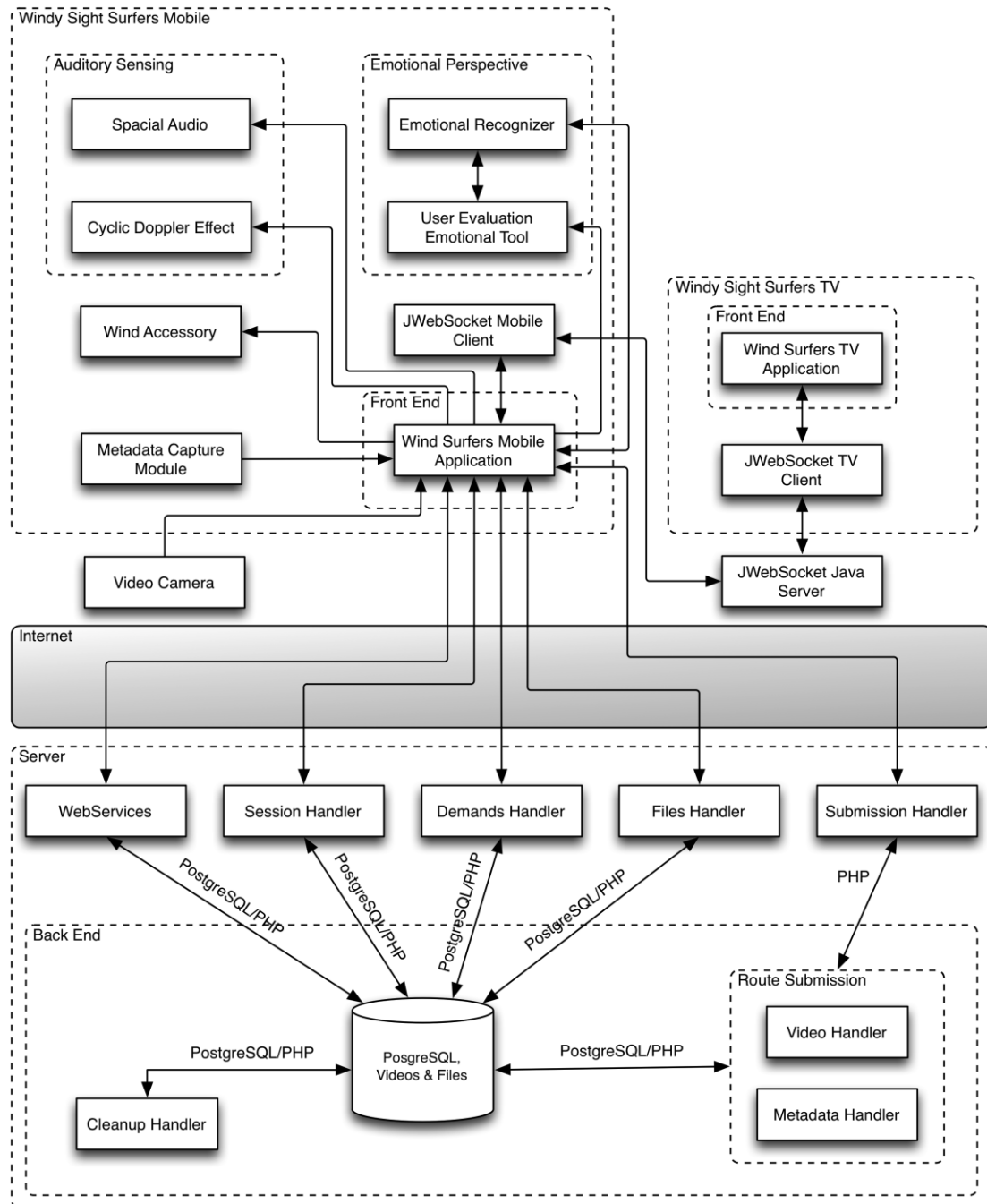


Figure 4.1: Windy Sight Surfers conceptual architecture

As it can be observed in Figure 4.1, the Windy Sight Surfers' architecture is structured around three major components: the Backend, the Windy Sight Surfers Mobile and the Windy Sight Surfers TV.

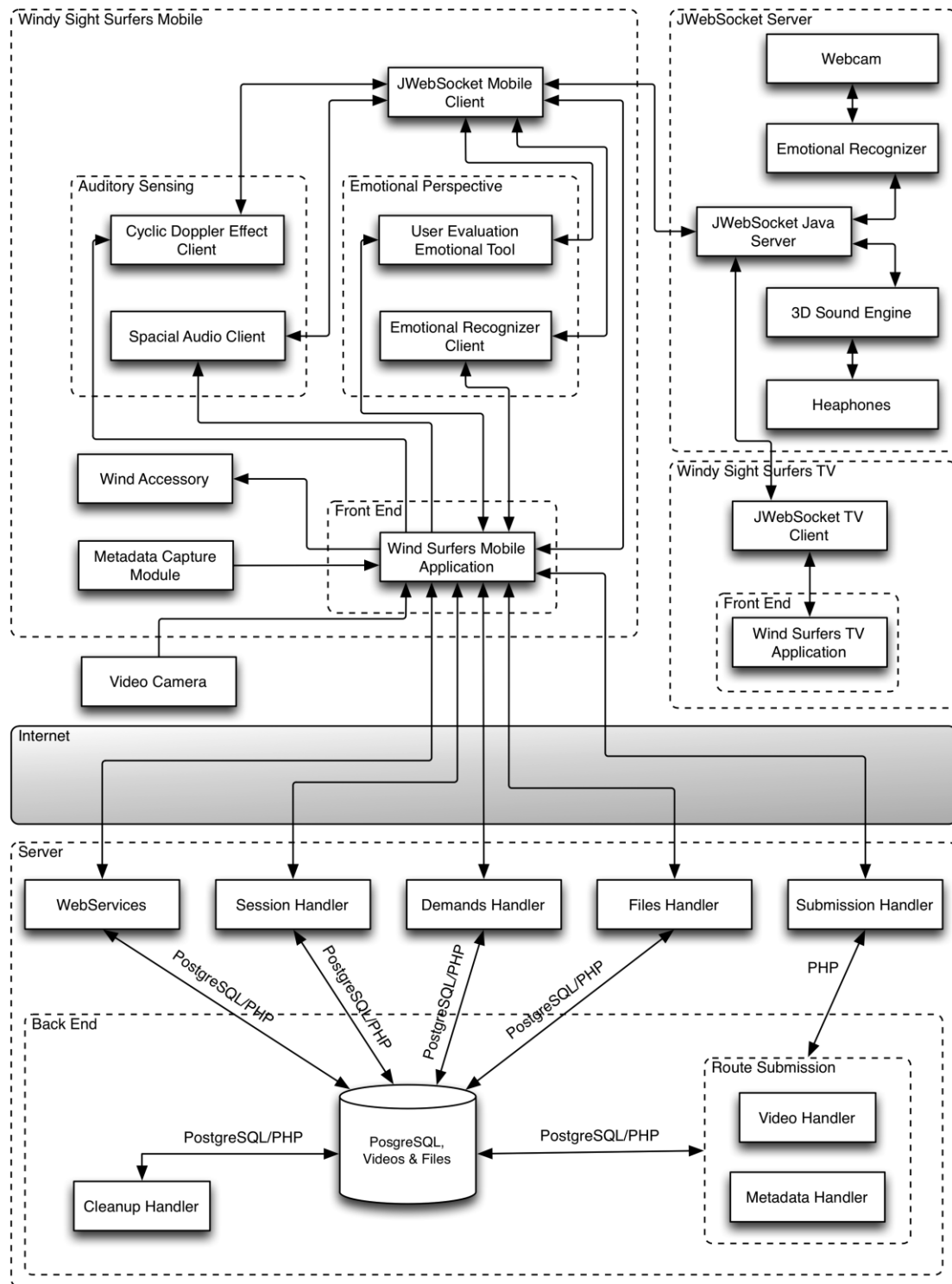


Figure 4.2: Windy Sight Surfers concrete architecture

The Backend component is responsible for the system's relational database and the several web services that allow frontend applications to access the information they need. The Windy Sight Surfers Mobile component comprises the mobile application and all the designed and developed modules related to it. Namely, this component includes the Video's Metadata Capture and Publishing components, the Perceptual

Sensing and the Context Awareness components, and the Emotional Perspective component (sections 3.3 to 3.8). Moreover, this component also comprises the second screen interface for the interaction with TV's, which are controlled by the Windy Sight Surfers TV component. With respect to the Windy Sight Surfers TV component, it is responsible for the mobile application extension through the interaction with TV's & Wider Screens.

Although Figure 4.1 highlights the conceptual architecture, several implementation restraints, which will be thoroughly described between sections 4.3 and 4.7, led to the result of slightly different system architecture (Figure 4.2).

As Figure 4.2 shows, the resulting architecture contains a fourth component (JWebSocket Server). The JWebSocket component also takes part in the conceptual architecture (figure 4.1), but played a simpler role, where it was responsible solely for the communication between the mobile and the TV components. It consisted of a Java server running JWebSocket (url-JWebSocket), which is a pure Java/JavaScript high-speed bidirectional communication solution for the Web. HTML5 is introducing a large amount of new technologies. The WebSocket protocol is a flexible and high-speed bidirectional TCP socket communication technology introduced in HTML5, which intends to replace the existing XHR approaches as well as Comet services. JWebSocket is an open source Java and JavaScript implementation of the HTML5 WebSocket protocol.

As it can be observed in Figure 4.2, the JWebSocket Server component actually became responsible for several of the modules that were supposedly under the control of the Windy Sight Surfers Mobile component. In the resulting architecture, the Auditory Sensing and Emotional Perspective modules became part of the JWebSocket Server component. These differences between the conceptual and the resulting architectures was due to several hardware barriers that were faced during system implementation, which will be explained between sections 4.3 and 4.7, as well as each of the referred components.

Considering the system's Backend, it is implemented using the PHP MVC (Model-View-Controller) framework CodeIgniter (url-CodeIgniter), which provides a very complete set of libraries to handle the HTTP requests and the database connection for transactions. Data requests are based on Object Relational Mapping (ORM), a software that allows a much more readable set of data. Considering that this system has a strong requirement regarding scalability, the MySQL relational database system (url-MySQL) was used as the system's DBMS. Aside from the database management system, the backend comprises a Submission module, a Session Controller module and a set of web services related to video/metadata transmission/reproduction, which are described next.

The Submission Handler is responsible for handling georeferenced videos' submissions to the server. Each time a georeferenced video is to be submitted, the Submission Handler starts a new session, which ends as the submission was completed. This handler comprises the Video Handler, which is responsible for receiving, validating and storing the video files in the server (updating the database using the Nginx software (url-Nginx) along with its *upload* and *upload-progress* modules), and the Metadata Handler, which is responsible for receiving, validating and distribute the metadata files.

The Cleanup Handler is a tool that executes on a regular schedule (using the Unix Crontab) to clean incomplete route submissions, which are considered failed submissions after an extended period of time (12 hours in this concrete case). It is necessary as the route submission is made in two phases: the video and metadata files are dependent on each other to complete the submission. Thus, the Cleanup Handler prevents the accumulation of incomplete and unnecessary data on server side.

As the user must be registered in the system in order to be able to submit a video to Windy Sight Surfers, the Session Handler is the responsible for the validation and authentication of users trying to log in the Windy Sight Surfers application, creating persistent sessions that last until the user logs out.

The Demands Handler answers the users' requests for video visualizations, opening a new unicast channel for each video request through the FFmpeg-server framework (url-FFmpeg). Along with the Demands Handler, the Files Handler is responsible for handling the access to all the files that are included dynamically in the application, being that front-end requests point to the file's unique key, and the server response is achieved through bitstream.

As a means to interact with the front-end applications, the backend has several web services available. Amongst their purposes, the search functionalities can be highlighted, as these offer tools to search content by different perspectives, including the search for videos according to a geographic region, which is used for efficiency purposes when the user is viewing the search results in the map (explained more thoroughly in section 4.4).

4.2 System Development Methodology

As the work of this thesis is part of large research project, it is intended to provide grounds for further research. Thus, it was necessary to adopt a development methodology whose outcome resulted in a cohesive system, where the design rationale supports a structure that easily accommodates change, which is inherent in projects with an iterative nature, and ease the inclusion of new components. The Model-View-

Controller (MVC) design pattern, which separates the modelling of the domain, the presentation, and the actions based on user input into three separate classes, fulfils these requirements. Moreover, it provides several other advantages. Namely, it is regarded for the way it eases the process of development and inclusion of new interfaces, by supporting multiple views. Also, testability is greatly enhanced when the MVC design pattern is employed. Testing components becomes difficult when they are highly interdependent, especially with user interface components. By separating the concern of storing, displaying, and updating information into three components that can be tested individually, MVC creates a much more appropriate testing approach, which is vital to the development of research applications.

Another important factor to choose the MVC design pattern is the fact that the core of this thesis is closely associated with the Android platform, which inherently uses this design pattern as the basic architecture for most of its applications.

In the MVC pattern, the View is the presentational aspect of an application that the user views and interacts with. In the Android platform, the View component is represented by Layouts, which are designed in the Graphical Layout Editor and are encoded in XML format.

The Model is the source of the information. In the Windy Sight Surfers, the model is composed by the Backend database component (described in section 4.1).

The backend data will be modelled as objects in code that can be manipulated in the Android application. In the Android platform, Activities are the ones responsible for getting the data, manipulate it, and then control how it is displayed in the layout of the application. Therefore, an Activity represents the Controller component, which is the main logic of the application.

4.3 Metadata Capture

This interface, as well as all the other interfaces of this application that make use of a map system, benefits from the Google Maps Android API v2 (url-GoogleMapsAndroidAPI). One of the main advantages of the version 2 of the Google Maps Android API relates to the fact that maps are now encapsulated in the MapFragment class, which is an extension of Android's Fragment class. This means it is now possible to implement Google Maps by extending the Android standard Activity class, rather than extending the MapActivity used in version 1, which makes Google Maps a much more flexible API.

The application's functionality for capturing the video's metadata can be broken down into four logical components: the GPS Tracking component, the Orientation

Tracking component, the Weather Status Tracking component, and the G-Force Tracking component. Except for the Weather Status Tracking component, which executes right at the beginning of the metadata capture process, all the components execute concurrently during the metadata capture process in separated *threads* of execution. When the metadata capture process starts executing, a parser tree is created using Java's DOM Parser (url-JavaDOM). The data resulting from the components' threads is added to the DOM document, and when the user indicates to stop the metadata capture process, this document is exported to an XML file, associated with the video, which contains all its metadata (Annex A).

4.3.1 GPS Tracking Component

This component is responsible to obtain the GPS coordinates associated to the route being recorded. Also, this component is responsible to obtain the current speed value of the device's movement. Therefore, a GPS connection is established by the application. Upon start, the application waits for a consistent GPS signal, which is considered to be obtained as soon as five GPS updates have been received.

Also, there might be periods of the recording where the GPS becomes unavailable, and the user must be informed of this. As the device might not be holding the device in his hand, the chosen alert method was the combination of a vibratory (the device vibrates for one second) and sonorous (short notification sound) alert. When this situation occurs, the GPS Tracking component still updates to the DOM document at the normal rate of half a second, but blank values are stored in it, which will be detected by the application when reading the metadata file. The outcome of this situation is that the application draws a straight line between every two coordinates, which were separated by a period in which the connection was lost.

4.3.2 Orientation Tracking Component

The Orientation Tracking component uses a digital compass as its basis, allowing the orientation information of the device to be captured and stored in the metadata file. For convenience while using the Windy Sight Surfers application, this component is designed in such a way the user can keep the mobile device in its pocket, or backpack, while filming the video. This is possible through an initial calibration, which resets the digital compass' orientation according to the actual position of the device. Also, the component contains a high-pass filter, which filters the sensor data, in order to discard negligible values that correspond to the natural user's movements, such as leg and arm swinging while walking.

In order to calibrate the device, the user must first select the “Calibrate the Device” option from the “Options” menu in the Capture Mode interface (described in section 3.3), and put the device in the position where it will be during the recording period. Then the device will collect orientation coordinates for a selected period of time between 10 and 60 seconds (selectable by the user), being that, the longer the collecting time, the better the calibration precision. When this period ends, the standard deviation of the acquired set of values is calculated using the following relationship:

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

$$\theta = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{X})^2}$$

Having the standard deviation value been calculated, the arithmetic mean of the values contained within standard deviation limits is calculated. This concludes the calibration process and provides the adjustment value, which is then applied to all the orientation values obtained by the orientation component.

With regards to the high-pass filter, it consists of a filter that attenuates the orientation sensor’s sensitivity. For the context of this work, a change in direction was established to be relevant if its angle is of at least 45°, which means the filter dismisses angle changes of smaller angles in order to accommodate to small involuntary movements and bumps.

4.3.3 Weather Status Tracking Component

When the user starts the metadata capture process, before the other three components of the metadata capture process start executing, the application uses a 3G Internet connection to collect weather information from the OpenWeatherMap’s web service (url-OpenWeatherMap). OpenWeatherMap is a weather forecast web service, which provides weather data in any location on the earth, as its current weather data is updated online based on data from more than 40,000 weather stations.

An example of a request to the OpenWeatherMap’s web service looks like the following:

```
"http://www.openweathermap.org/data/2.1/find/city?lat=" + lat + "&lon=" + lon + "&cnt=1"
```

Listing 4.1: OpenWeatherMap’s web service request

The first part of it corresponds to the link of the web service itself, “lat” and “lon” represent the latitude and longitude, respectively, values of the place where the user is standing with the device at the time (Obtained through the GPS Tracking component), and “&cnt=1” indicates to the web server that the request is only interested in receiving

data from the most significant station (the one that is located closer to the real location). The last part of the request is needed because, if not specified, the answer from the web service would retrieve information relative to several stations, which is not desired.

The web service response is in the JSON (url-JSON) format and looks like the following:

```
{
  "message": 0.0137,
  "cod": "200",
  "calctime": "",
  "cnt": 1,
  "list": [
    {
      "id": 2265726,
      "name": "Moscavide",
      "coord": {
        "lon": -9.10222,
        "lat": 38.779289
      },
      "distance": 1.051,
      "main": {
        "temp": 287.73,
        "humidity": 78,
        "pressure": 1010,
        "temp_min": 285.93,
        "temp_max": 288.71
      },
      "dt": 1370398834,
      "wind": {
        "speed": 0.51,
        "deg": 308
      },
      "clouds": {
        "all": 0
      },
      "weather": [
        {
          "id": 800,
          "main": "Clear",
          "description": "Sky is Clear",
          "icon": "01n"
        }
      ]
    }
  ]
}
```

Listing 4.2: OpenWeatherMap’s web service response

This response contains several data values that will be stored in the metadata file being constructed at this time. The Google Gson Java library (url-GSON) was used to convert the JSON responses into Java Objects, which in its turn, were converted in instances of the DOM document used during the metadata capture process. More specifically, the values corresponding to the following fields are retained: “name”, “temperature” (in Kelvin, converted to Celsius by subtracting 273.15 from the given

value), “wind speed” (in mps), “wind deg” (orientation in degrees), and “weather id” (according to the OpenWeatherMap weather condition codes list ([url-OpenWeatherMapWeatherCodes](http://openweathermap.org/weather-conditions))).

If the user does not have an Internet connection available at the moment of the recording, blank values are stored in the DOM document and consequently in the metadata file, which the user can edit later through the “edit” option from the publishing menu (described in section 3.4).

4.3.4 G-Force Tracking Component

During the metadata capture process, every half second, the Android’s accelerometer is used to capture and calculate the G-Force values. With an accelerometer is possible to obtain an array of values that indicate the acceleration applied to a device, including the force of gravity, in units of $1 = 1G = 9.8 \text{ m/s}^2$ along the X, Y, and Z axes (Figure 4.3). When the device is lying flat on a horizontal surface in front of the user, the X axis goes from left to right, the Y axis goes from the user toward the device, and the Z axis goes upwards perpendicular to the surface.

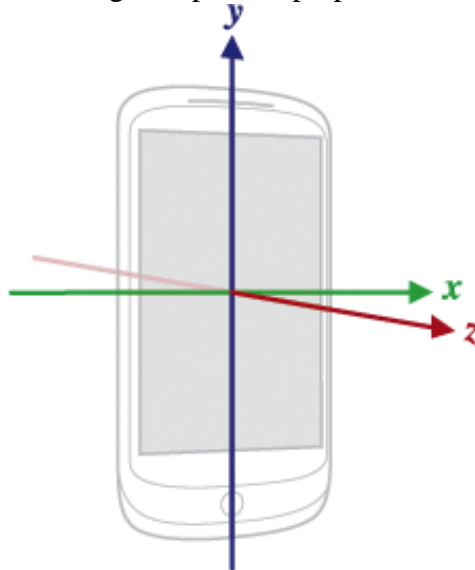


Figure 4.3: Accelerometer’s Axis

Conceptually, an accelerometer determines the acceleration (in G’s) that is applied to a device (A_d) by measuring the forces that are applied to the sensor itself (F_s) using the following relationship:

$$A_d = -\sum F_s / \text{mass}$$

However, the force of gravity (G) is always influencing the measured acceleration according to the following relationship:

$$A_d = -G - \sum F_s / \text{mass}$$

For this reason, when the device is sitting stationary (and not accelerating), the accelerometer reads a magnitude of $1G = 9.81 \text{ m/s}^2$. Similarly, when the device is in free fall and therefore rapidly accelerating toward the ground at 9.81 m/s^2 , its accelerometer reads a magnitude of $g = 0 \text{ m/s}^2$. This means that, in order to measure the real acceleration of the device, the contribution of the force of gravity must be removed from the accelerometer data. This was achieved by applying a high-pass filter. Conversely, a low-pass filter was used to isolate the force of gravity (which is needed in order to apply the high-pass filter).

For the desired purposes of the application, the graphical display of the G-Force values during video reproduction is done using two of its dimensions (Figure 3.19), as users are familiarized with this graphical representation from other applications. More specifically, the X and Y axes were used for the graphical representation. As for the value below the graph, it represents the overall G-Force value (considering the 3 dimensions). Listing 4.3 presents a code snippet describing how this was implemented.

```
public void onSensorChanged(SensorEvent event){
    // alpha is calculated as t / (t + dT),
    // where t is the low-pass filter's time-constant and
    // dT is the event delivery rate.
    double alpha = 0.8;
    double gravity[] = new double[3];
    double linear_acceleration[] = new double[3];

    // Isolate the force of gravity with the low-pass filter.
    // The resulting array contains the G-Force values
    // in the 3 dimensions.
    gravity[0] = alpha * gravity[0] + (1 - alpha) * event.values[0];
    gravity[1] = alpha * gravity[1] + (1 - alpha) * event.values[1];
    gravity[2] = alpha * gravity[2] + (1 - alpha) * event.values[2];

    // Remove the gravity contribution with the high-pass filter.
    linear_acceleration[0] = event.values[0] - gravity[0];
    linear_acceleration[1] = event.values[1] - gravity[1];
    linear_acceleration[2] = event.values[2] - gravity[2];

    // Obtaining the G-Force value based on the three dimensions
    // given by the accelerometer.
    double globalGForce = Math.sqrt(
        linear_acceleration[0] * linear_acceleration[0]
        + linear_acceleration[1] * linear_acceleration[1]
        + linear_acceleration[2] * linear_acceleration[2]);
}
```

Listing 4.3: G-Force calculation through the Android's Accelerometer

4.4 Location Based Video Search

When a user conducts a search, either through the keywords and filters system or through the map, it is converted in a standard MySQL query and sent to the database, following the standard search procedures.

The search results depend on the overload of information. That is, in situations where too many results for a small geographic area are associated to the conducted search, a clustering technique is applied. In these cases, the application groups the results in several clusters, which are presented in the map as bubbles (being each bubble's size directly related to its number of elements). As the user zooms in and out the results are automatically grouped/ungrouped. The way the system generates the bubbles is based on the division of the map viewport in a number of bounding boxes, which is constant regardless of the zoom level. In order to avoid bubble overlapping, which might occur when adjacent bounding boxes contain a high number of videos, the algorithm has a second phase, where the map viewport is analysed and cases where adjacent bounding boxes contain a high number of videos are detected and treated, being that in each of these cases, the adjacent bounding boxes are joint, forming one larger bounding box. Instead of executing just once, the number of times the second phase of the algorithm executes is relative to the zoom level, as in lower zoom levels it is more likely that bubble overlapping occurs. As in the android Google Maps Android API the zoom level ranges between 2 and 21, the algorithm's second phase executes once if the zoom level is between 15 and 21, twice if it is between 10 and 15, three times if it is between 5 and 10, and four times if it is between 2 and 5 ([url-GoogleMapsAndroidAPIZoomLevels](#)). It is important to mention that bubble overlapping might still occur, as this algorithm does not completely prevent overlapping from occurring, but its occurrence chances decrease drastically, and its presence is not noticeable to user.

When the viewport focuses on a map area that (at the selected zoom level) contains a little enough number of videos for the algorithm to consider that video overload is not present, the clusters (represented by bubbles) are replaced by the real videos, represented in the map as routes, as previously described.

A problem that may occur and is not solved at the time of this writing (and this might be considering for future work) relates to the fact that, in certain zones the population of the system with videos is expected to be much higher when compared to other zones. As an example, the accumulation of videos in Central Europe or East Asia is expected to be much higher than in Western Asia or Africa. These discrepancies are related with several factors, such as the population density and the development index in the different areas. This problem may have as a result the presentation of obvious results to the user (regarding the bubble system), as the bubbles may be located around big cities. In order to solve this problem the bubble-drawing algorithm is to consider these differences and adjust the bubble size to these factors. These adjustments consider examples such as the fact that 10 videos in a small village in Africa might be more significant than 500 videos in London.

4.5 Perceptual Sensing Features

With regards to the Perceptual Sensing features of the application, several approaches were considered, and components relative to them were designed and implemented. More precisely, the designed perceptual sensing features focus on visual, tactile, and auditory approaches as a means to increase immersion. This section describes the implementation aspects and problems faced during the implementation process.

4.5.1 Visual Sensing in 360° Video

As it was described in section 3.6.1, 360° videos are mapped onto a transitional canvas that is in turn rendered around a cylinder, to represent the 360° view and allow the feeling of being “partially” surrounded by the video. As the Android standard video library, the `VideoView` class (`url-VideoView`), did not offer tools to implement this functionality, it was required to use the `SurfaceView` class (`url-SurfaceView`). The `SurfaceView` class provides a dedicated drawing surface embedded inside of a view hierarchy. Therefore, the format and size of this surface are controllable. This was also used to move the video’s position to the left or right on the canvas, in order for the various angles of the 360° video to appear on the screen.

In order to allow users to continuously pan around the 360° video in both left and right directions by moving the tablet around, as if the tablet was a window to the 360° video surrounding the user, the first implementation approach was to use the digital compass of the Android device to detect orientation changes. Although this was a plausible solution, some important flaws supported the choice of another approach. More specifically, the use of the Digital Compass alone, can led to inaccuracies in this process, because when digital compasses turn around themselves multiple times, drift occur, which, after a while using the device, translates into inaccuracies ranging from 5° to 20° on each circumference drawn by the user while holding the device. As the functionality serves precisely this purpose, this error is not tolerable. Another problem is related to the fact that, using the Digital Compass alone, the user cannot turn the device over the Z axis for more than 90°. In other words, if the user is using the device in parallel to his eyes in a seated position and changes to a lying position while retaining the device parallel to his eyes, the orientation angle will clip by 180°.

In the second approach to this problem, sensor fusion was used, which consists of using the gyroscope, accelerometer and magnetic field sensors within the tablet device in conjunction. Alongside, a filter that eliminates the gyroscope drift and the signal noise of the accelerometer and magnetometer were also combined, thus allowing measuring the device’s rotations accurately.

Also, a high-pass filter was added to this functionality, whose function is discarding sensor data that is related to the user's involuntary movements. In other words, this filter is needed in order to decrease the sensors sensitivity, by only considering movements that correspond to the user's intention to move the tablet right or left.

4.5.2 Wind Accessory

The Wind Accessory is a prototype based on the Arduino Mega ADK open-source electronics prototyping platform ([url-ArduinoMegaADK](#)) and its source code was written in the Arduino's variation of the C programming language. The device was designed to be mounted on the back of the tablet and it controls two fans generating a combined air flow of 180CFM (when operating at maximum RPM). Using an USB connection, the Wind Accessory is connected to the Android device and they exchange messages through the Android Open Accessory (AOA) protocol ([url-AndroidAOA](#)). Having the conceptual and algorithmic aspects of the Wind Accessory been described in section 3.6.2, this section focuses on its implementation process and hardware architecture.

The integration of the Android and Arduino platforms faced some problems, mainly due to the fact that, at the date of this writing, the integration platforms (such as the Android Open Accessory Protocol) are still at an embryonic stage.

The Android SDK organizes Android applications in four distinct components: Activities, Content Providers, Services and Broadcast Receivers. An Activity is an application component with which users can interact (through the screen), being that each activity is given a window in which to draw its user interface. An Android application consists of multiple activities that are loosely bound to each other. Figure 4.4 represents an Android Activity's lifecycle.

The `onPause()` method comes into play whenever another activity comes into the foreground. This method created integration problems, because when returning control to the activity that controls the Wind Accessory, which is achieved through the `onResume()` method, another instance of the communication between the two devices is created, and repeated/inconsistent messages start to be exchanged between both devices. Adding an ID to each communication session and considering only the messages with the latest ID (on the Arduino side) could easily solve this problem, but this would dramatically increase the number of exchanged messages while using the accessory, and ultimately would create a battery drain on the Android device. Therefore, the `onPause()` and `onResume()` methods were overridden in order to destroy the current communication session when the `onPause()` method is called, and establish a new

communication session each time the `onResume()` method is called. This required to change the structure of the standard Android class `UsbManager`.

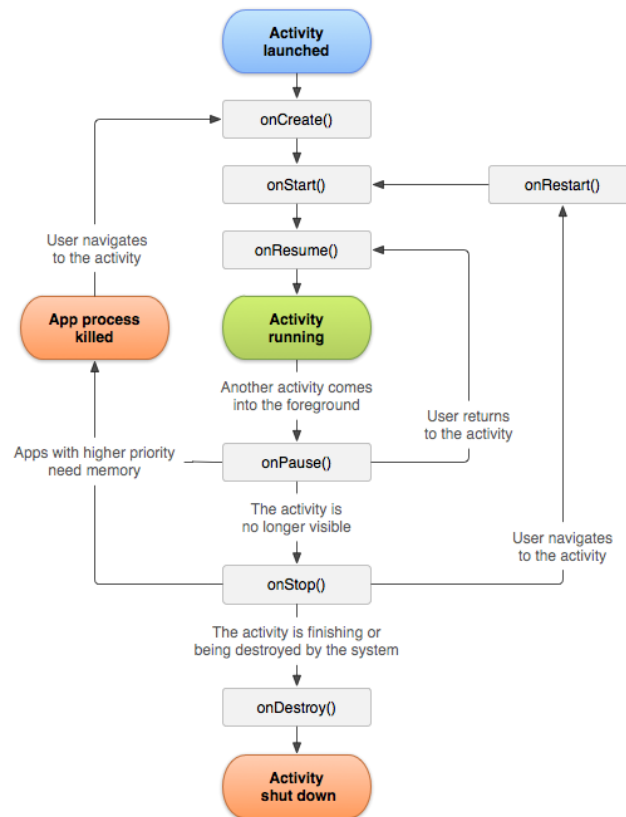


Figure 4.4: Android Activity's lifecycle

Another problem surfaced when it was detected that opening a second instance of the Windy Sight Surfers application without deleting the first instance from memory, creates a communication malfunction, as the first instance of the application is frequently stored in memory, and thus cached as a process. When this situation occurs, while launching the second instance, the first instance can be called from the cache, which will result in an exception, when trying to establish a connection with the Wind Accessory. The implemented solution Focused on the Android application to identify if there is any open connection before it tries to establish a new one, and if the response is positive, the Android application asks the Wind Accessory for the connection information (ID and used communication ports), which upon retrieval enable the normal communication.

Regarding the Wind Accessory hardware architecture, it consists of a box that contains the required control electronics, two fans, and an USB cable, which is used to connect to the Android device (Figure 4.5).

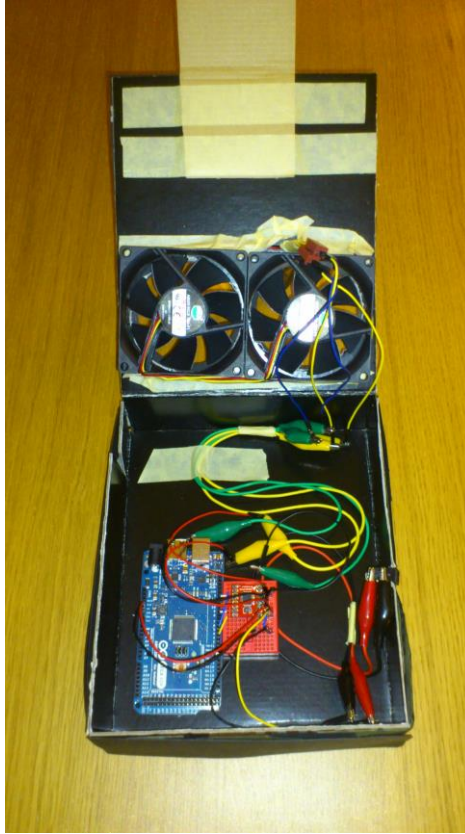


Figure 4.5: Wind Accessory inside view

The Android device is mounted on the front of the Wind Accessory through a frame on the front of the box of the accessory (Figure 4.6).



Figure 4.6: Wind Accessory inside view

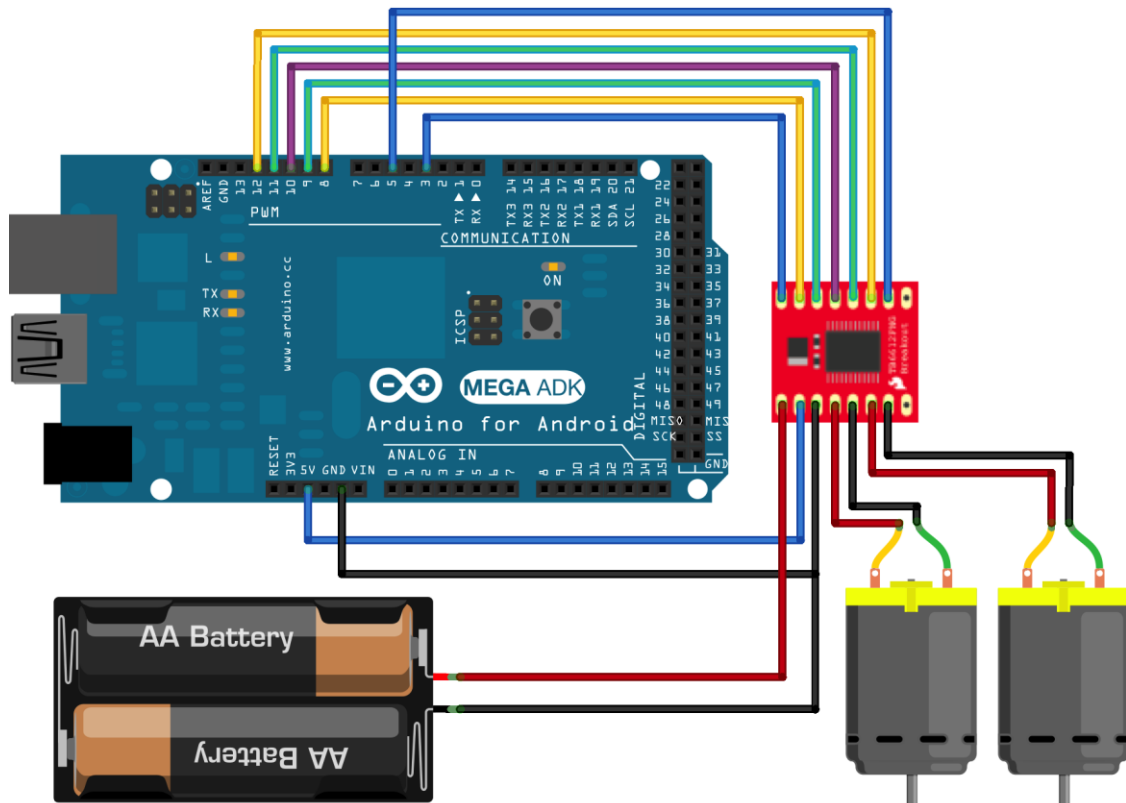


Figure 4.7: Wind Accessory's architecture (breadboard perspective)

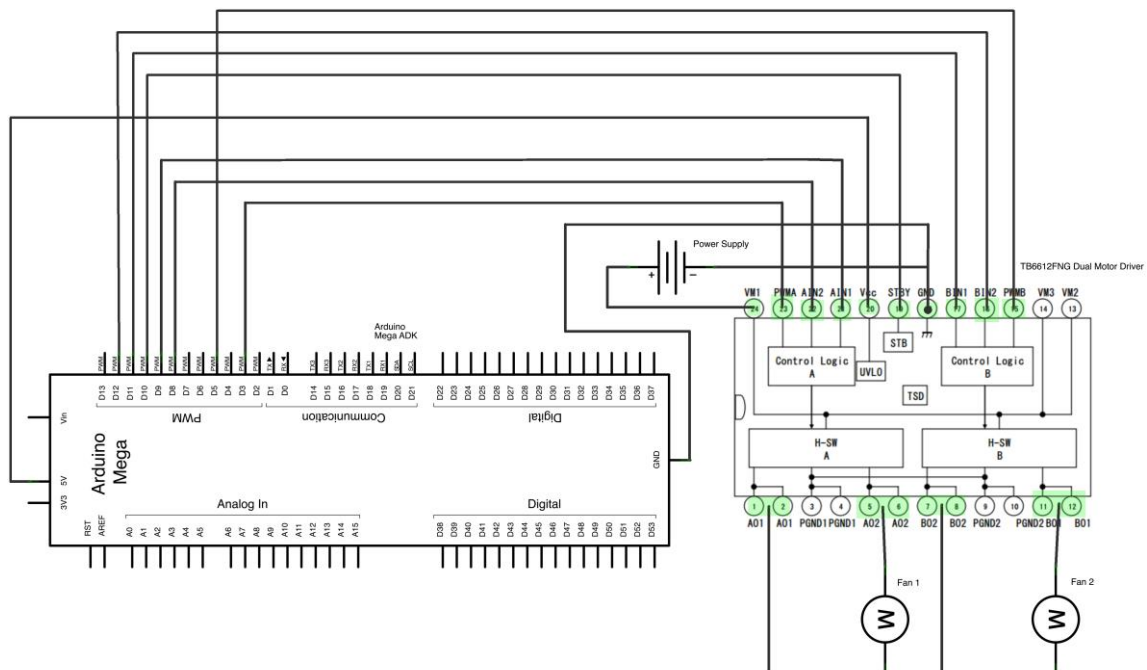


Figure 4.8: Wind Accessory's architecture (schematic perspective)

The Arduino is powered at 5V, but the used fans are powered at 12V. Therefore, in order to power both fans, a dual Motor Driver was used (in electronic terms, a fan is analogous to a simple motor). Namely, the SparkFun TB6612FNG Dual Motor Driver (url-SparkFunDualMotorDriver) was used. In order to power the motor driver, a 12V

transformer can be used (for testing purposes). But, as this application is targeted at mobile environments, an option was also implemented, which allows the motors to be powered through a pair of 6V batteries (the process is conceptually analogous to the 12V transformer option). Figures 4.7 and 4.8 illustrate the Wind Accessory's architecture through the breadboard and schematic perspectives, respectively.

The way the fans are operated in the Arduino platform is through Pulse Width Modulation (PWM), which consists in a technique for getting analog results with digital means. This technique uses digital control to create a square wave (a signal switched between on and off). This on-off pattern can simulate voltages in between full on (5 Volts) and off (0 Volts) by changing the portion of the time the signal spends on versus the time that the signal spends off. The duration of "on time" is called the pulse width. By changing, or modulating, that pulse width one gets varying analog values, which is exactly what is needed to operate the fans at controlled speeds. In Figure 4.9, the green lines represent a regular time period. In Arduino, a call to `analogWrite()` is on a scale of 0 - 255, such that `analogWrite(255)` requests a 100% duty cycle (always on), and `analogWrite(127)` is a 50% duty cycle (on half the time) for example. When translating this to the real effect on the controlled fans, the call `analogWrite(127)` results in the fans operating at half speed, whereas the call `analogWrite(255)` results in the fans operating at full speed.

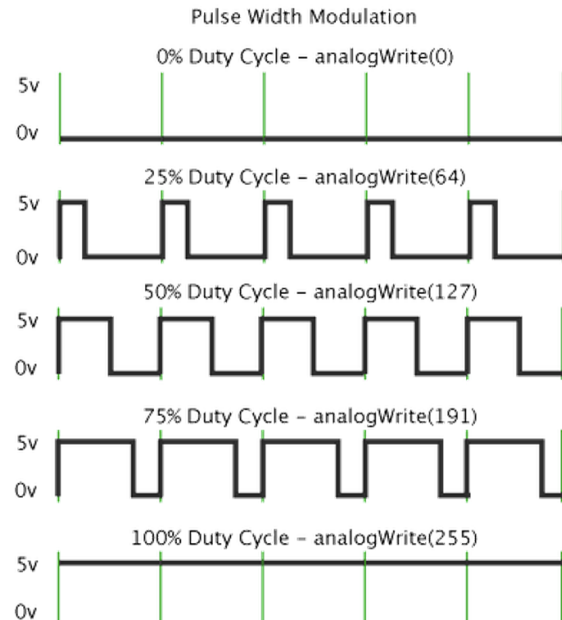


Figure 4.9: Pulse Width Modulation

In its original design, the device had the fans perpendicularly angled to the user. However, as it will be described in section 5 (Evaluation), initial tests revealed that users disliked that characteristic, because the wind was being blown into their eyes. This happened because users naturally adjust the tablet (and the Wind Accessory by

consequence) in order to get the best viewing angle, which translates in a perpendicular alignment between the users' eyes and the screen. In order to solve this problem, the device's design was revised, and the fans were fixed at a slightly lower angle in the device (approximately 15°), pointing to the area between the user's nose and neck.

4.5.3 Auditory Sensing

Regarding the audio component of the Windy Sight Surfers application, the 3D audio space was implemented using the Web Audio JavaScript API ([url-WebAudioAPI](#)), which means any standard set of stereo headphones can reproduce the created effect, although high quality headphones are able increase the realism of the referred effect. Web Audio API enables audio processing and synthesizing in web applications. It features several capabilities found in modern game audio engines and some of the mixing, processing, and filtering tasks that are found in modern desktop audio production applications. Using this library, it is possible to select the coordinates of multiple sound sources in a 3D environment, being that the user's ears are located in the center of the 3D space. Conceptually, the space can be thought of as a three dimensional cube, where its center represents the user's ears. One can then select any coordinate in the “(x,y,z)” form for each of the sound sources.

In reality this feature was not implemented in the Android platform, but rather it was designed and implemented on a computer in JavaScript (running in a standard web browser). This was due to the fact that, at the time of this writing, there are very few 3D audio options available on the Android platform, and the ones that exist are very limited in what they offer. Plus, they all required the application to use the Android's low level Native Development Kit ([url-AndroidNDK](#)), which was not desired for the context of the Windy Sight Surfers application. On the other hand, JavaScript's Web Audio API represents a very accessible and powerful audio tool, and once the communication between a standard computer and an Android device would be necessary for the “Interaction with TV's & Wider Screens” component of this thesis (section 3.9), it was decided to profit from these resources and implement this component through the Web Audio API. Therefore, once the communication between the mobile application's Auditory Sensing component and the JWebSocket Server's 3D Sound Engine component is established through the JWebSocket Java Server (described in section 4.1), the audio features will execute upon command from the mobile application.

Still it is important to note that that could be a more attractive alternative to this implementation approach. Java Android applications profit from the WebView class, which consists on a view that displays web pages. Although this seems a much more attractive alternative, the performance of this approach reaches very low levels (extended sound latency). However, this is purely a hardware barrier, and once it is

transposed, the component can quickly be moved to a WebView inside the Windy Sight Surfers application (as it is implemented in JavaScript, which is a cross-platform language).

The Cyclic Doppler Effect feature was also implemented using this approach. Conceptually, the Doppler Effect can be described as the change in the observed frequency of a wave, occurring when the source and/or observer are in motion relatively to each other. These changes in the frequency of a wave are relative, not consisting of concrete frequency variations from the sound source. As an example, let's consider the situation of a car moving at a constant speed, and during its course, it approaches, passes and recedes an observer. As the car is moving at a constant speed, the frequency wave of the sound of its engine does not change. Still, the observer notices a very clear change in the observed frequency of a wave. Namely, when the car approaches the observer, the frequency increases, and when the car recedes from the observer, the frequency decreases. This effect occurs because, when the car is approaching the observer, each successive wave cycle is emitted from a closer position to the observer than the previous wave cycle. This creates a situation where each emitted wave takes less time than its predecessor to reach the observer, and thus increases the frequency. The inverse process is analogous.

Sound waves need a medium to be propagated. By consequence, the observer and source's velocities are relative to the medium in which the waves are transmitted. This translates to the fact that the total Doppler Effect may result from motion of any of these three components (source, motion of the observer and motion of the medium).

The relationship between observed frequency f and emitted frequency f_0 is given by the following relationship:

$$f = \left(\frac{c + v_o}{c + v_s} \right) f_0$$

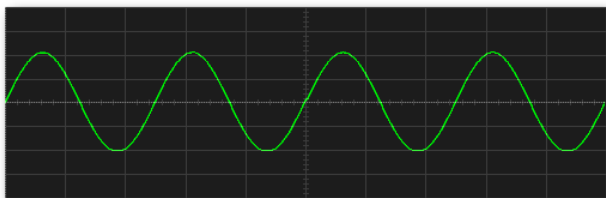
Where c represents the waves' velocity in the medium;

v_o represents the velocity of the observer (positive if the distance between the observer and the source is decreasing, otherwise negative);

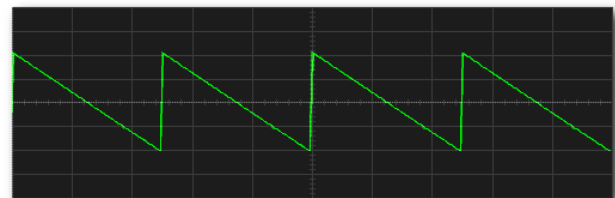
v_s represents the velocity of the source (positive if the distance between the observer and the source is increasing, otherwise negative).

In this functionality, the main objective was to investigate if the use of the Doppler Effect cyclically can increase the movement sensation when viewing videos. When implementing this component, a strong emphasis was put on the sound source's nature. Different sounds can drastically change the impact of this feature. Therefore, different sounds, with different characteristics were experimented.

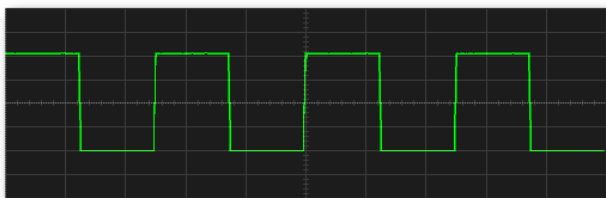
The first sound to be experimented was the sound of a car's engine. This is one of the sounds that may create a stronger movement sensation in the user. However, due to the fact that, from a conceptual point of view, an engine sound is very complex (it does not consist of any waveform but rather consists of the combination of a large amount of waveforms), it can also be one of the most intrusive sounds. Therefore, several other sounds were experimented. Low frequency sounds (sounds with a low pitch) are known to be less intrusive than other sounds. This led to the experimentation of different sounds that reflect these characteristics. In order to create these sounds, an Analog synthesiser was used (url-Moog). The sounds were recorded in a computer directly running the synthesiser's output through an audio interface. The recorded and experimented sounds were created with the intention to reproduce the four main sound waveforms and investigate which, given their simplicity, are more suitable for the purpose of this application. Namely, it was designed one sound based on each of the following waveforms:



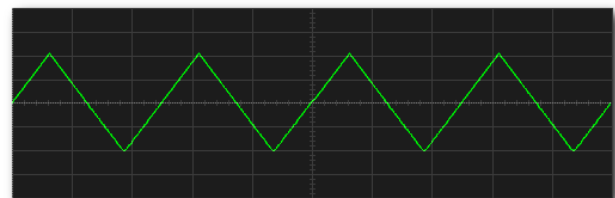
a) Sine Wave



b) Sawtooth Wave



c) Square Wave



d) Triangle Wave

Figure 4.10: Waveforms used in sound design

The referred sounds were designed through an oscillator, which consists of a repeating waveform with a fundamental frequency, peak amplitude, and associated shape. As it has been referred, the low frequency sounds are known to be less intrusive. Therefore, the designed sounds are based on low frequencies. Also, the threshold and optimal volumes of the Doppler Effect sound layer were measured, thus establishing the peak amplitude value of the designed sounds. In this way, this experiment focused on the variation of the shape of the waveforms, being a distinct shape associated with each sound.

Figure 4.10a describes a Sine Wave, which is often considered the most fundamental building block of sound. As the sine wave stands as the purest waveform,

any other waveform can be created through the sum of a series of sine waves. As the name suggests, this sound wave is based on the trigonometric Sine function

Figure 4.10b highlights a Sawtooth Wave. Sawtooth waves are characterised by having a strong, clear, buzzing sound. One way to generate a sawtooth wave is by adding a series of sine waves with different frequencies and volume levels (amplitudes). The first sine wave is the loudest and its frequency corresponds to the heard frequency of the resulting sawtooth, and is thereby referred to as the fundamental frequency. Each one of the other sine waves that make up a sawtooth is progressively quieter, and they have frequencies which are integer multiples of the fundamental frequency (each sine wave corresponds to a different multiple). These frequencies are referred to as harmonics.

Figure 4.10c depicts a Square Wave. Despite having a rich sound with a bright and rich timbre, square waves' sound can be described by being not quite as buzzy as a sawtooth wave, but not as pure as a sine wave. Like sawtooth waves, it is possible to generate square waves by adding a series of sine waves with decreasing volume levels. The characteristic that defines square waves contains is that they contain only the odd numbered harmonics.

Figure 4.10d depicts a Triangle Wave. Triangle waves sound can be described as something between a sine wave and a square wave, as triangle waves have a softer timbre when compared to square or sawtooth waves. Like square waves, they contain only the odd harmonics of the fundamental frequency. However, they differ from square waves in the sense the volume of each added harmonic drops faster. Also, despite sharing several geometric similarities with the sawtooth wave, triangle waves have two sloping line segments.

4.6 Emotional Perspective

As it happened with the Auditory Sensing component (described in section 4.5.3), in reality the emotional perspective features were not implemented in the Android platform, but rather they were designed and implemented on a computer in the C++ programming language. This was due to the fact that, at the time of this writing, the available emotion recognizer frameworks based on facial expression recognition target mainly desktop systems, and the vast majority of them are proprietary systems, with very expensive usage licenses. On the other side, under the scope of the ImTV project, the team from the Universidade Nova de Lisboa – Faculdade de Ciências e Tecnologia is developing an emotion recognizer framework (through facial expressions recognition), which ended up being the perfect fit for the desired application. This framework is in the C++ programming language and runs on a standard computer using

a webcam to recognize emotions out of facial expressions. In the Windy Sight Surfers system, the emotional features relied on this framework, which was incorporated in the JWebSocket Server (described in section 4.1), and it is called from the JWebSocket Java Server component through the Java Native Interface (JNI), which is an interface that enables Java code to interoperate with applications written in other programming languages such as C and C++ (url-JavaNativeInterface). In this context, and as this was the best available option, instead of using the front-facing camera of the mobile device, a standard webcam was used (url-LogitechWebcam), being that it was connected to the computer running the Emotional Recognizer component and coupled with the mobile device as a substitute for the device's front-facing camera. Therefore, once the communication between the mobile application's Emotional Perspective component and the Emotional Recognizer component is established through the JWebSocket Java Server, the emotional perspective features will execute upon command from the mobile application.

4.7 Interaction with TV's & Wider Screens

As it was described in section 3.9, a widescreen computer screen was used to simulate a TV. Therefore, the computer is running the web application, which based on HTML5 and is compliant with any standard web browser supporting WebGL (url-WebGL). When it starts executing, the web application establishes a WebSocket connection with the JWebSocket Server (described in section 4.1), which persists while the application is running. Once the connection with the JWebSocket Server is established, the web application does not execute any "visible" action until the Windy Sight Surfers mobile application also establishes a connection with the JWebSocket Server. However, until a mobile device is connected, the web application is cyclically sending broadcast UDP packets, indicating to possible mobile devices running the Windy Sight Surfers application that there is a "TV" nearby to which they can be connected, and expand the application. As the mobile application is "listening", when such message is detected, the user is given the possibility to extend the application to the TV screen. As the user accepts the proposal, the mobile application establishes a WebSocket connection with the JWebSocket Server, and this connects the two applications to one another. Once connected, the mobile application rearranges itself, being that the TV becomes responsible for reproducing the videos, while the mobile application is responsible for the control of the TV, and it displays additional metadata. The following sections describe some of the second screen's functionalities in more detail.

4.7.1 Controlling the TV's Video Viewing Angle Through the Mobile Application

In the second screen context, the minimap drag interface is the main tool to control the viewing angle (pan around the 360° video). Through this interface, the user may change the video angle by simply pressing at point or drag to any point in the minimap, being that the selected minimap point by the user becomes the center of the video's viewport.

Alternatively, users can pan around the 360° video by using the device as if it was a steering wheel. When holding the device and in turning it to the left or right, the video viewing angle is shifted to the left or right, respectively. Similarly to the Visual Sensing feature (section 4.5.1), a high-pass filter was needed for this feature, whose function is discarding sensor data that is related to the user's involuntary movements. The filter decreases the sensors sensitivity, by only considering movements that correspond to the user's intention to turn the tablet right or left.

4.7.2 Geographical Navigation & Orientation in the 360° Videos and Maps

One of the functionalities of the mobile application as a second screen relates to the ability to present a map, which identifies video trajectories, including the trajectory associated to the video being video in the TV screen.

As it was described in section 4.3.1, the video's metadata file contains its geographical information, which describes the route associated to the video. These geographical coordinates are captured by the GPS at a constant rate (every half second), and each of this coordinates is associated to a timestamp. Therefore, by summing the video's current reproduction time to the initial geographical coordinate's timestamp, it is always possible to determine the geographical point associated with the current time of the video being viewed. In the context of this functionality, this information is used to move a marker on the map accordingly to the video's current playtime, as the following pseudocode describes:

```
1. while video.isPlaying == true
2. loop
    3. currentMetaTime = initialMetaTime + video.currentReprodTime
    4. currentMetaLat = metafile.getLat(currentMetaTime)
    5. currentMetaLon = metafile.getLon(currentMetaTime)
    6. mapView.setVideoMarker(currentMetaLat, currentMetaLon)
    7. mapView.centerMap(currentMetaLat, currentMetaLon)
8. end loop
```

The referred dynamic marker highlights the geographical point associated with the current playtime, and contains an interface similar to the Video View Area feature, which indicates the map area related to the angle of the video being viewed at each moment. In order to do so, this feature takes the orientation information in the video's metadata file into consideration. One important characteristic of this feature is that updates to the marker's Video View Area are animated, because if they reported plainly the orientation information contained in the video's metadata, the marker would produce abrupt changes, which would confuse users. Therefore, these changes were animated so that they resemble a continuous change. Overall, the marker's Video View Area updating process follows the following pseudocode:

1. **while** video.isPlaying == true
2. **loop**
 3. currentMetaTime = initialMetaTime + video.currentReprodTime
 4. currentMetaOrientation = metafile.getOrientation(currentMetaTime)
 5. videoMarker.updateViewAreaAnimated(currentMetaOrientation)
6. **end loop**

At any time during video reproduction, the user can press and drag the video marker to any point of the route in the map, resulting in a video playback time jump to the selected route position. Moreover, as the map illustrates other routes that were recorded in the map area contained in the screen, the video marker may as well be moved to any point of these routes, resulting in the change of the video to the time associated to the selected route position.

During video reproduction, intersections with other videos that cross the current video (geographically) are indicated in the video through hyperlinks and a vibrating alert (in the case the hyperlink position is located outside the viewport currently being viewed on the TV screen). Such hyperlinks are associated to a specific video and geographical point. Therefore, when a user selects one of these hyperlinks, the associated video starts reproducing from the associated geographical point.

Chapter 5

User Evaluation

This section describes the user evaluation that the Windy Sight Surfers application went through. The main objective of the Windy Sight Surfers' user evaluation was to investigate whether and in which conditions the designed features contribute to a more immersive environment for the user in accordance with the research questions defined in section 1.2.

Being the first Research Question (RQ1) defined as “Do the designed map search features enhance the search process?”, tests focused on analysing the benefits the map component introduces in the search process. Considering “Would a full screen pan-around interface increase the sense of immersion ‘inside’ the 360° video?” as the second Research Question (RQ2), tests were conducted to see what is the perceived immersion benefit of the use of the feature that allows using the tablet as a window to the 360° video when compared to the “pan around” drag interface. With regards to the Wind Accessory, focus shifted to the Research Question “Does wind contribute to increasing realism of sensing speed and direction in video viewing?” (RQ3). Its contribute to the realism of the sense of speed in video viewing was analysed, being that attention was given not only to the speed sensation, but also to the orientation factor. As for the spatial sound feature, the Research Question “Does a 3D mapping of the video sound allow for easier identification of the video orientation while it is being reproduced?” (RQ4) was addressed. Tests were carried out to verify the benefits of this feature to the orientation of the user while viewing 360° video. Different values for several characteristics of the 3D sound environment were manipulated in order to find the values that create an ideal sound environment. About the Cyclic Doppler Effect feature, the Research Question “Can a controlled use of the Doppler Effect increase the movement sensation while viewing videos?” (RQ5) was considered. The conducted experiments focused on finding the most suitable sound and volume level for this feature, as well as on identifying which are the situations that can profit from this feature, which is a crucial aspect of this feature, as if it is not applied to the right situations it can become intrusive. With regards to the designed context awareness

features, focus shifted to the Research Question “Do the designed context awareness features contribute to a more immersive environment?” (RQ6). Their immersive contribution was evaluated. Also, the importance of taking into account the user’s preferences when considering which of the available information must and must not be presented to the user was analysed. Regarding the Emotional Perspective features, and considering the current state of this module, the analysis focused on the Research Question “Do users consider the emotional perspective relevant in the access and search of videos?” (RQ7). Considering the Research Question “Does the interaction with TVs & wider screens, with video in full screen and additional content and navigation control in a second screen, contribute to a more immersive environment?” (RQ8), the features related with the interaction with TV’s and wider screens features were analysed in terms of their ease of use and intuitiveness and contribution to a more immersive environment.

As well as testing each of the designed features individually, the global immersive capabilities of the environment were considered during user evaluation. The following sections present the method of this evaluation, its results, and lastly, a final overview of the results and the major conclusions that can be drawn from the user evaluation.

5.1 Method

The user evaluation was performed on two occasions during the development of this work, and different modules were evaluated at each one. This was mainly due to distinctive phases corresponding to the middle and end of the project. Results were also reported in two scientific papers that were submitted and accepted for publication, as described in section 1.4. The first evaluation considered only the Visual Sensing (section 3.6.1), Tactile Sensing (section 3.6.2), Context Awareness (section 3.7) and the use of Second Screens (section 3.9) modules. In the second, and final evaluation, all the designed modules were tested, including the ones tested during evaluation one, which made it possible to gather new information to confirm or update the results of evaluation one, now that the system is in a more complete state. Below, the evaluation method that was applied in both evaluation one and two is described.

Using the usability dimensions proposed in the USE questionnaire (Lund, 2001), each feature’s perceived Usefulness, Satisfaction and Ease of use was analysed. In order to evaluate dimensions related to Immersion, the Self-Assessment Manikin (SAM) (Lang, 1980; Suk, 2006; url-SAM) was used, coupled with additional parameters of Presence and Realism (PR). The Self-Assessment Manikin is a non-verbal pictorial assessment technique that directly measures the pleasure, arousal, and dominance associated with a person’s affective reaction to a specified stimulus (Figure 5.1). In

other words, SAM measures emotion, which is closely associated with immersion. Also, during evaluation two, the designed emotional evaluation tool (described in section 3.7.3) was used during the evaluation to recognize the most prevalent emotions. Lastly, the global immersiveness in the Windy Sight Surfers was evaluated.

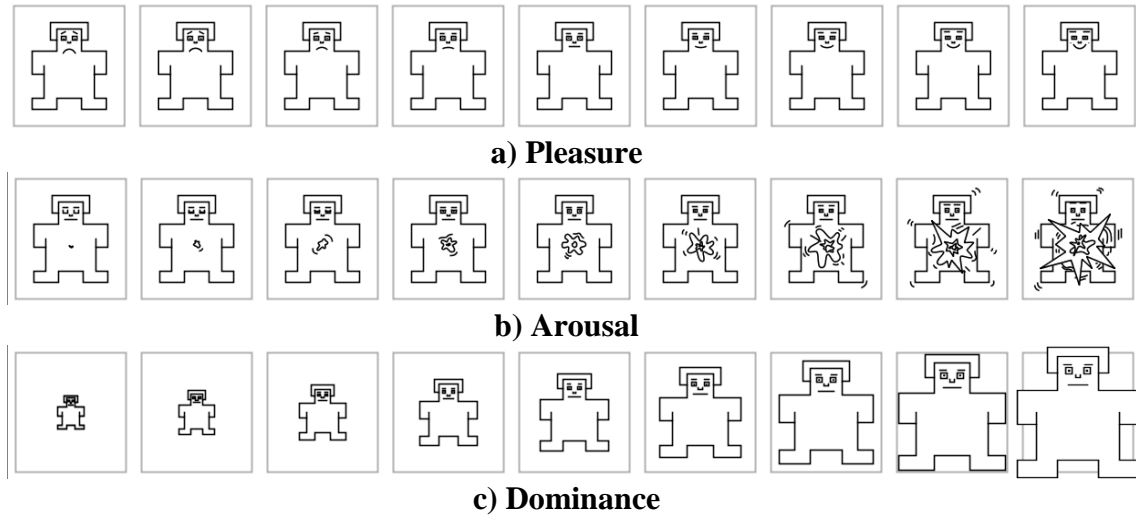


Figure 5.1: Self-Assessment Manikin 9-Point Scale

A task-oriented evaluation of Windy Sight Surfers was performed based mainly on Observation, Questionnaires and semi-structured Interviews. After explaining the purpose of the evaluation and giving a short briefing about the concept behind Windy Sight Surfers, demographic questions were asked, followed by a task-oriented activity. Errors, hesitations and user performance were observed and annotated.

At the end of each of the tasks, users provided a 1-5 USE rating, a 1-9 SAM rating and a 1-9 PR rating regarding the functionality focused by the task. Comments and suggestions were annotated after each task was completed. Also, in evaluation two, during each of the tasks, the most prevalent emotions identified by the emotion recognizer were logged. Annexes B and C contain the scripts for user evaluation's one and two, respectively. These scripts were filled by the evaluator, being that all the questions referred in them were asked verbally to the user, except in the SAM rating case, where users were shown the set of three images (depicted in Figure 5.1) and asked to choose a value from each of them, thus providing the SAM rating. At the end of the session, users were asked to rate the overall application.

Once Immersion is a central topic of this thesis, special attention was given to the evaluation of the Immersive capabilities of the created environment. Slater states that Presence is a human reaction to Immersion (Slater et al, 2009). Therefore, by evaluating presence, one can tell about immersion capabilities of the system. To do so, users completed an adapted version of the seven-point scale format Immersive Tendencies Questionnaire (ITQ) (Witmer & Singer, 1998) before the experiment (Annex D), and an

adapted version of the Presence Questionnaire (PQ) (Witmer & Singer, 1998) after the experiment (Annex E), with 31 questions each, also in a seven-point scale format.

The user evaluation population consisted of 21 individuals (8 female, 13 male), aged between 18 and 57 years old. This group of individuals was the same for evaluation one and two. In terms of literacy, all users had at least finished high school, they were all familiar with the concept of accessing videos on the Internet, but only 5 had previously interacted with 360° videos, and only 6 had heard about 3D audio. Regarding Evaluation One, the foreseen time for the completion of the 11 tasks was 30 minutes. As for Evaluation Two, the foreseen time for the completion of the 34 tasks was 80 minutes. Considering the extended duration of evaluation two, and taking into account possible user fatigue, a small break of 5 minutes was given to each user at half time during the process (40 minutes). In both evaluations, all users met the foreseen time for the tasks' completion, and they detected the new functionality underlying each new task they were asked to perform.

5.2 Results

Results are divided in nine subsections, concerning: the Video Search features, the Perceptual Sensing features, the Context Awareness features, the Perceptual Sensing & Context Awareness features in Conjunction, the Emotional Perspective features and the Interaction with TV's & Wider Screens Features. For each of these subsections, results are commented along the corresponding tasks and features, and tables are presented highlighting the Mean and Std. Deviation for USE, SAM and PR. In these tables, values related with evaluation one and evaluation two are presented side by side (in the cases of features that were evaluated in both evaluations one and two). Afterwards, a subsection presents the emotional impact of Windy Sight Surfers (summarized in Table 5.17), followed by the Immersive Tendencies and the Presence Questionnaires results (summarized in Tables 5.18 and 5.19). Lastly, a subsection presents global overall comments. Each task is given an abbreviation name (e.g. "T1"), where the number relates to the task's number in evaluation two, and therefore, range from T1 to T34.

5.2.1 Video Search Features

Regarding the several search mechanisms designed, tests were carried out to analyse their benefits in the search process. Namely, users had to complete tasks where they were supposed to conduct a video search of a specified video based on the set of filters provided by the application and on a specified set of keywords (T1); conduct a video search of a specified video using the "search through map" feature (T2); find and select the specified video through the cover-flow method (T3); and find and select the specified video through the map (T4). Users showed positive signs of surprise when

first having contact with the search methods related to the map, often stating that they had never seen a search mechanism based on geolocation, and that it can be very useful for certain types of searches. In addition, most users pointed out the need for both the conventional and the map search approaches, referring that it is nice to have the possibility to choose between them, as each of the approaches may be more appropriated for certain searches, thus answering RQ1 by indicating that the proposed map search features enhance the search process.

Features in Task:	Usefulness		Satisfaction		Ease of Use	
	M	σ	M	σ	M	σ
T1 Search through filters & keywords	4.4	0.5	4.6	0.3	4.4	0.4
T2 Search through map	4.5	0.5	4.3	0.6	4.1	0.6
T3 Find video through the cover-flow	4.5	0.3	4.5	0.4	4.7	0.3
T4 Find video through the map	4.6	0.5	4.5	0.4	4.3	0.7

Table 5.1: USE evaluation of the Video Search features (scale 1-5)

Features in Task:	SAM						PR			
	Pleasure		Arousal		Dominance		Presence		Realism	
	M	σ	M	σ	M	σ	M	σ	M	σ
T1 Search through filters & keywords	8.0	0.5	7.2	0.9	8.8	0.2	6.6	1.3	7.0	1.1
T2 Search through map	8.4	0.9	8.3	0.6	8.0	0.8	7.5	1.4	7.3	0.9
T3 Find video through the cover-flow	8.2	0.4	8.0	0.5	8.7	0.2	6.7	1.2	6.9	1.4
T4 Find video through the map	7.8	0.8	8.1	0.5	7.8	1.1	7.1	1.5	7.0	0.8

Table 5.2: SAM and PR evaluation of the Video Search features (scale: 1-9)

5.2.2 Perceptual Sensing Features

Users were asked to: move around a 360° video by moving the tablet around (using only this method) (T5); move around a 360° video by using the drag interface (using only this method) (T6); view a 360° video with the Wind Accessory, where they were asked to try and move the tablet around (paying attention to the wind) in order to identify the wind direction (T7). Users appreciated the tested features, especially the video navigation by moving the tablet around and the wind accessory. Regarding the video navigation by moving the tablet around, users reported to be a more natural approach when compared to the touch interface, thus contributing to a more immersive experience. However, the consensus among users was that, although the “tablet moving” feature creates a more realistic experience, there are situations where the drag interface can be more suitable, giving examples such as while they are viewing a video

while being seated on the bus. This result answers RQ2, reinforcing the idea that both interfaces are needed and complement each other. Also, between evaluation one and evaluation two, some improvements were made to the sensitivity of the “tablet moving” feature, which are reflected in the better results of evaluation two, especially in the Satisfaction and Dominance fields, where users stated they felt more in control of the environment.

Features in Task:	Usefulness		Satisfaction		Ease of Use	
	M	σ	M	σ	M	σ
T5 View video moving the tablet around	<i>4.8</i>	<i>0.4</i>	<i>4.5</i>	<i>0.5</i>	<i>4.8</i>	<i>0.4</i>
	4.8	0.3	4.7	0.3	4.8	0.5
T6 View video using the drag interface	<i>4.8</i>	<i>0.4</i>	<i>4.5</i>	<i>0.5</i>	<i>4.8</i>	<i>0.4</i>
	4.8	0.4	4.5	0.5	4.8	0.4
T7 Wind Accessory	<i>4.3</i>	<i>0.6</i>	<i>4.6</i>	<i>0.8</i>	<i>4.9</i>	<i>0.3</i>
	4.3	0.6	4.8	0.7	4.9	0.3

Table 5.3: USE evaluation of Perceptual Sensing features – Visual and Tactile (scale 1-5), (evaluation one in italics, evaluation two in bold)

Features in Task:	SAM						PR			
	Pleasure		Arousal		Dominance		Presence		Realism	
	M	σ	M	σ	M	σ	M	σ	M	σ
T5 View video moving the tablet around	<i>8.1</i>	<i>0.6</i>	<i>8.7</i>	<i>0.8</i>	<i>7.7</i>	<i>0.6</i>	<i>8.6</i>	<i>0.9</i>	<i>8.6</i>	<i>0.7</i>
	8.2	0.5	8.7	0.8	8.0	0.7	8.8	0.8	8.8	0.7
T6 View video using the drag interface	<i>8.1</i>	<i>0.6</i>	<i>8.7</i>	<i>0.8</i>	<i>7.7</i>	<i>0.6</i>	<i>8.6</i>	<i>0.9</i>	<i>8.6</i>	<i>0.7</i>
	7.9	0.5	7.4	0.8	8.8	0.5	8.3	0.8	8.4	0.7
T7 Wind Accessory	<i>8.2</i>	<i>0.9</i>	<i>8.6</i>	<i>0.9</i>	<i>8.6</i>	<i>0.8</i>	<i>8.9</i>	<i>0.5</i>	<i>8.9</i>	<i>0.5</i>
	8.3	0.8	8.6	0.9	8.6	0.7	8.9	0.4	8.9	0.5

Table 5.4: SAM and PR evaluation of Perceptual Sensing features – Visual and Tactile (scale: 1-9), (evaluation one in italics, evaluation two in bold)

Relating to the wind accessory, users pointed out that it allowed a more realistic sense of speed in video viewing, as the PR results show (T7: PR: 8.9; 8.9), thus confirming RQ3. Identifying the wind direction was relatively straightforward, being that only two users seemed a bit confused at first, and the reason was the same for both of them: when they were trying to identify the wind direction, the video they were viewing was in a situation where it changed direction rather frequently, which created a situation where the wind direction changed also rather frequently. Once the video direction stabilized, those users had no problem identifying the wind direction. However, several users pointed out that the device had the fans perpendicularly angled

to the user, which created a situation where the wind was being blown into their eyes. This led to a revision in the Wind Accessory's design, where the angle of the fans was adjusted, as it was described in section 4.3.3.2. Although one cannot infer that it was a cause-effect relation, it is fairly safe to assume that this adjustment to the fans eliminated the problem pointed out by several users during evaluation one, as in evaluation two no user pointed out to this specific point and the Satisfaction and Pleasure values registered slight improvements.

In order to test the Spatial Audio feature, users were asked to: view a 360° video with headphones being that the video's sound was standard stereo sound (T8); and view the same video with the 3D sound capability (T9). Regarding T9, users were asked to vary the virtual "distance" of the simulated speakers to the users' head through the manipulation of a seekbar (the seekbar value was relative to the radius of the virtual circle associated with the distance of the sound sources to the user's head), and find the optimal virtual distance between users' head and the sound sources. In respect to RQ4, users stated that this feature provided them with a better sense of orientation, and they preferred the 3D sound version, with the restriction that the sound source must be located between 1 and 3 meters (in the virtual sound space) of the users' head.

Features in Task:	Usefulness		Satisfaction		Ease of Use	
	M	σ	M	σ	M	σ
T8 Stereo sound	4.7	0.7	4.5	0.7	4.8	0.5
T9 3D sound	4.8	1.1	4.7	1.0	4.8	0.5

Table 5.5: USE evaluation of Perceptual Sensing features – Spatial Audio (scale 1-5)

Features in Task:	SAM						PR			
	Pleasure		Arousal		Dominance		Presence		Realism	
	M	σ	M	σ	M	σ	M	σ	M	σ
T8 Stereo sound	7.9	0.7	8.1	0.7	8.8	0.7	8.5	0.9	8.4	0.8
T9 3D sound	8.3	1.2	8.5	1.3	8.6	1.0	8.7	1.0	8.8	0.9

Table 5.6: SAM and PR evaluation of Perceptual Sensing features – Spatial Audio (scale: 1-9)

In order to test the Cyclic Doppler Effect feature, users were asked to view videos with the Doppler Effect feature activated, being that sounds with different characteristics were used in each video to create the Doppler Effect. In the first video the sound of a car's engine was used (T10). In the second video, the created and recorded low frequency sound waves (described in section 4.3.3.3) were used (T11). Users needed to choose their preferred sound being that, in T11 users were asked to

vary the Doppler Effect sound by choosing their preferred of the four created sounds from a radio button group. Also users were asked to vary the sound volume through a seekbar and identify the optimal value for the Cyclic Doppler Effect feature. Next, users viewed three videos with the Doppler Effect feature activated and calibrated to the user's preferences (indicated in T10 and T11), and were asked to state in which of them they liked the Doppler Effect the most. The three videos presented situations where the degree of movement was: 1) high (T12); 2) medium (T13); and 3) little movement (T14). The order in which the videos were viewed was randomized for each of the users. Lastly, taking into consideration all the user's preferences regarding the Doppler Effect feature (indicated in T10, T11, T12, T13 and T14), users viewed a video two times: one without (T15) and the other with the "custom" Doppler Effect feature (T16), and were asked to state whether they felt the Doppler Effect feature increased the movement sensation or not.

Answering RQ5, users declared that the Doppler Effect feature increased the movement sensation, which is supported by the SAM and PR values (T16: SAM: 8.8; 8.6; 8.7; PR: 8.9; 9), although its reproduction must be carefully controlled. Namely, all users preferred one of the low frequency sounds. More specifically, the sound consisting of a sine waveform was the most popular of the four designed sounds (15 out of 17 users selected the sine waveform), with users stating that it was the smoothest sound. The remaining two users selected the sound consisting of a triangular wave form, which conceptually is a middle point between a sine wave and a square wave, which means that, excluding the sine wave, it is the wave form which contains the less amount of harmonics, thus meaning it is the second smoothest sound. The main conclusion regarding the sound type is that users consider the smoothest sounds to be the most adequate for this purpose. When comparing the designed sounds with the car's engine sound, all users referred to the designed sounds as much more unobtrusive sound, underlining their adequacy to the situation. With regards to the volume, users tended to set the Doppler Effect volume level between 7% and 18% of the main video sound volume. When discussing which are the situations that can profit the most from this feature, according to the users' feedback, the more movement there is in a video, the more satisfying the Doppler Effect Becomes. Users referred to T12 (high degree of movement) to express a situation where they particularly enjoyed the Doppler Effect (T12: USE: 4.8; 4.7; 4.9; PR: 8.8; 8.9). On the other side, results shown that in videos with little/no movement, the viewing experience is better without this feature, as the USE and PR results show (T14: USE: 2.8; 3.1; 4.9; PR: 3.9; 3.3). This result confirmed there is a minimum amount of movement required, and led to the development of a high-pass filter that added the requirement for a minimum amount of movement in order for the Doppler Effect simulation to execute. When viewing the video with all the

preferences adjusted, all users declared that the Doppler Effect feature increased the movement sensation, which is supported by the SAM and PR values (T16: SAM: 8.8; 8.6; 8.7; PR: 8.9; 9) and answers RQ5.

Features in Task:	Usefulness		Satisfaction		Ease of Use	
	M	σ	M	σ	M	σ
T10 Car's engine sound	2.7	0.6	2.5	0.5	4.9	0.2
T11 Designed low frequency sounds	4.4	0.9	4.6	0.8	4.9	0.2
T12 High degree of movement	4.8	0.4	4.7	0.4	4.9	0.2
T13 Medium degree of movement	4.4	0.4	4.5	0.4	4.9	0.1
T14 Little degree of movement	2.8	0.4	3.1	0.4	4.9	0.1
T15 Without the Doppler Effect	4.7	0.5	4.5	0.6	4.9	0.2
T16 With the custom Doppler Effect	4.7	0.6	4.7	0.6	4.9	0.1

Table 5.7: USE evaluation of Perceptual Sensing features – Cyclic Doppler Effect (scale 1-5)

Features in Task:	SAM						PR			
	Pleasure		Arousal		Dominance		Presence		Realism	
	M	σ	M	σ	M	σ	M	σ	M	σ
T10 Car's engine sound	3.9	1.0	6.1	0.8	7.4	1.2	4.0	1.6	3.7	1.4
T11 Designed low frequency sounds	8.2	0.9	7.9	0.8	8.4	1.0	8.2	0.6	8.2	0.5
T12 High degree of movement	8.7	1.0	8.7	0.5	8.5	0.2	8.8	0.5	8.9	0.6
T13 Medium degree of movement	8.0	1.7	7.8	0.8	8.1	0.6	8.0	1.8	7.9	1.5
T14 Little degree of movement	4.4	0.7	5.2	0.5	6.2	0.4	3.9	0.6	3.3	0.8
T15 Without the Doppler Effect	7.6	1.0	7.7	1.3	8.6	1.2	8.3	1.1	8.4	1.0
T16 With the custom Doppler Effect	8.8	1.2	8.6	0.8	8.7	0.9	8.9	1.0	9.0	0.7

Table 5.8: SAM and PR evaluation of Perceptual Sensing features – Cyclic Doppler Effect (scale: 1-9)

5.2.3 Context Awareness Features

Regarding the context awareness features, users were asked to: view a video with the overlay, albeit with only the permanent information activated, and identify each category of information provided (T17); view a predefined set of three videos about a specific theme with the overlay that has both the permanent and momentary information activated, where users were intended to identify the permanent category related to each example of momentary information when there was a relation, being that all the

momentary information available was always presented (T18); and (T19), similar to T18, except that the momentary information presented took into account the user preferences (users were given the opportunity to choose which of the momentary information features were activated). Some users stated that, when it did not take into consideration the users' preferences, some of the information being presented momentarily were not appreciated, problem that was solved with the mechanism tested in T19. This result means that taking user's preferences into account when considering which of the available information must and must not be shown is of great importance. After that, users were asked to, while watching a 360° video, find and follow a link to a crossing trajectory (T20). Users could easily find and follow the links.

Features in Task:	Usefulness		Satisfaction		Ease of Use	
	M	σ	M	σ	M	σ
T17 Only the permanent information	<i>4.3</i>	<i>0.7</i>	<i>4.3</i>	<i>0.8</i>	<i>4.6</i>	<i>0.5</i>
	4.3	0.7	4.3	0.7	4.6	0.2
T18 Permanent and momentary information	<i>2.9</i>	<i>1.3</i>	<i>2.8</i>	<i>1.0</i>	<i>4.6</i>	<i>0.5</i>
	3.1	1.1	3.0	1.0	4.4	0.4
T19 Permanent and Custom momentary information	<i>4.2</i>	<i>0.7</i>	<i>4.0</i>	<i>0.5</i>	<i>4.6</i>	<i>0.5</i>
	4.6	0.6	4.0	0.5	4.6	0.2
T20 Follow a link to a crossing trajectory	<i>4.2</i>	<i>0.9</i>	<i>4.2</i>	<i>0.8</i>	<i>4.9</i>	<i>0.3</i>
	4.9	0.4	4.8	0.4	4.9	0.2

Table 5.9: USE evaluation of Context Awareness features (scale 1-5), (evaluation one in italics, evaluation two in bold)

Features in Task:	SAM						PR			
	Pleasure		Arousal		Dominance		Presence		Realism	
	M	σ	M	σ	M	σ	M	σ	M	Σ
T17 Only the permanent information	<i>8.0</i>	<i>0.7</i>	<i>8.0</i>	<i>0.4</i>	<i>8.0</i>	<i>0.6</i>	<i>8.0</i>	<i>0.8</i>	<i>7.7</i>	<i>0.6</i>
	8.1	0.5	8.0	0.3	8.0	0.6	8.1	0.9	7.8	0.6
T18 Permanent and momentary information	<i>6.8</i>	<i>1.8</i>	<i>6.9</i>	<i>1.4</i>	<i>7.1</i>	<i>1.0</i>	<i>7.2</i>	<i>0.9</i>	<i>7.6</i>	<i>1.2</i>
	6.8	1.7	6.9	1.4	7.1	0.9	7.3	1.3	7.6	1.2
T19 Permanent and Custom momentary information	<i>8.3</i>	<i>0.8</i>	<i>8.3</i>	<i>0.9</i>	<i>8.3</i>	<i>0.4</i>	<i>8.3</i>	<i>0.4</i>	<i>8.4</i>	<i>0.6</i>
	8.5	0.6	8.3	0.7	8.4	0.5	8.3	0.4	8.4	0.7
T20 Follow a link to a crossing trajectory	<i>8.3</i>	<i>0.5</i>	<i>7.9</i>	<i>0.9</i>	<i>8.5</i>	<i>0.7</i>	<i>8.6</i>	<i>0.5</i>	<i>8.4</i>	<i>0.9</i>
	8.3	0.5	8.1	0.8	8.5	0.7	8.6	0.4	8.4	0.9

Table 5.10: SAM and PR evaluation of Context Awareness features (scale: 1-9), (evaluation one in italics, evaluation two in bold)

Answering RQ6, users stated that the context awareness features contribute to a more immersive environment (when taking user preferences into account) in the sense

they provide the viewer with additional information about the videos, which allow them to further connect to the video. Between evaluation one and two, some improvements were made to the design of the context awareness features. Namely, the initial analog speedometer was considered too intrusive by some users in evaluation one, which led to the design of a simpler digital speedometer (tested in evaluation two). These changes are reflected in the better results of evaluation two. Although there are not very significant differences in the evaluation results, a consistent improvement was verified. Especially, the Usefulness and Satisfaction fields register significant improvements in the four tasks related to the context awareness features.

5.2.4 Perceptual Sensing & Context Awareness features in Conjunction

The purpose of this evaluation was to try to identify which are the main benefits of the two categories of features. Users were asked to watch and interact with a video in four modes: 1) with all of the created features (the perceptual sensing and the context awareness features) activated (T21); 2) with the perceptual sensing features disabled (T22); 3) with the context awareness features disabled (T23); and 4) with all the features disabled (just bare video) (T24). The order in which these four tests were performed was randomized for each of the users. All users preferred the environment reproduced on T21, with all features activated. Furthermore, from Table 5.11, it can be observed that the inclusion of both categories of features does not add extra complexity to the environment (both T21 and T24 scored the same value in the “Ease of Use” parameter).

Features in Task:	Usefulness		Satisfaction		Ease of Use	
	M	σ	M	σ	M	σ
T21 All features	<i>4.9</i>	<i>0.4</i>	<i>4.9</i>	<i>0.4</i>	<i>4.9</i>	<i>0.4</i>
	4.9	0.3	4.9	0.4	4.9	0.2
T22 Perceptual sensing features disabled	<i>4.6</i>	<i>0.5</i>	<i>4.3</i>	<i>0.7</i>	<i>4.6</i>	<i>0.5</i>
	4.6	0.5	4.3	0.6	4.7	0.3
T23 Context awareness features disabled	<i>4.4</i>	<i>0.6</i>	<i>4.4</i>	<i>0.6</i>	<i>4.9</i>	<i>0.3</i>
	4.4	0.5	4.5	0.6	4.9	0.2
T24 All features disabled	<i>2.7</i>	<i>0.7</i>	<i>2.6</i>	<i>0.7</i>	<i>4.9</i>	<i>0.3</i>
	2.7	0.4	2.8	0.7	4.9	0.2

Table 5.11: USE evaluation of Perceptual Sensing & Context Awareness features in Conjunction (scale 1-5), (evaluation one in italics, evaluation two in bold)

Users expressed they do prefer perceptual sensing features in terms of satisfaction, and prefer context awareness features in terms of usefulness. This matches with the

categories of designed features. While the Perceptual Sensing features are more closely related to increase immersion by providing the user with a more realistic video environment, the Context Awareness features are able to increase immersion by making the user more aware of the characteristics of the environment. It is important to highlight that the results of evaluation two validate the results of evaluation one, as they were fairly similar.

Features in Task:	SAM						PR			
	Pleasure		Arousal		Dominance		Presence		Realism	
	M	σ	M	σ	M	σ	M	σ	M	σ
T21 All features	<i>8.6</i>	<i>0.5</i>	<i>8.7</i>	<i>0.4</i>	<i>8.5</i>	<i>0.3</i>	<i>8.8</i>	<i>0.3</i>	<i>8.9</i>	<i>0.2</i>
	8.7	0.3	8.7	0.2	8.5	0.3	8.8	0.3	8.9	0.2
T22 Perceptual sensing features disabled	<i>7.9</i>	<i>0.8</i>	<i>7.7</i>	<i>0.6</i>	<i>8.5</i>	<i>0.6</i>	<i>8.1</i>	<i>0.6</i>	<i>8.2</i>	<i>0.4</i>
	7.9	0.7	7.8	0.8	8.5	0.5	8.1	0.6	8.3	0.6
T23 Context awareness features disabled	<i>8.4</i>	<i>0.6</i>	<i>8.3</i>	<i>0.9</i>	<i>8.4</i>	<i>0.3</i>	<i>8.1</i>	<i>0.7</i>	<i>8.6</i>	<i>0.4</i>
	8.4	0.6	8.3	0.9	8.4	0.3	8.3	0.8	8.6	0.4
T24 All features disabled	<i>7.0</i>	<i>1.0</i>	<i>6.2</i>	<i>1.5</i>	<i>8.5</i>	<i>0.4</i>	<i>7.3</i>	<i>1.2</i>	<i>7.2</i>	<i>1.1</i>
	7.1	1.1	6.3	1.4	8.5	0.4	7.3	1.2	7.2	1.0

Table 5.12: SAM and PR evaluation of Perceptual Sensing & Context Awareness features in Conjunction (scale: 1-9), (evaluation one in italics, evaluation two in bold)

5.2.5 Emotional Features

In order to test the EmoMap feature, users were asked to: search a video in which the dominant emotion is anger through the EmoMap (T25); search a video in which the dominant emotion is happiness through the checkboxes filter tool (T26); search a video in which the dominant emotion is happiness by exemplifying the desired emotion to the camera (T27); search for neutral videos and view the results in the cover-flow mode (T28); search for neutral videos and view the results in the map (T29). All users highlighted the EmoMap was their preferred way to look for videos when the filter in use is emotion, qualifying it as intuitive and innovative. Regarding the search through the checkboxes filter mode and the emotion detection mode, in general users preferred the emotion detection mode, although four users preferred the checkboxes mode, stating it was just because, in this mode, they were sure the system was searching for what they were looking for. Despite the fact the majority of users did not feel this way, this can be an indicator that a small percentage of users still does not trust enough in this kind of systems, and prefer less intuitive, but more familiar methods to achieve their goals. With respect to the different ways of displaying the search's results, users enjoyed the

map (with the bubbles), stating that the search process was more useful when they had access to the geo-location variable.

When evaluating the EmoMe feature users were asked to identify what information was contained in each of the graphs and charts (T30); and choose and view a recommended video (T31). All graphs and charts were correctly identified and users easily understood the recommendations system.

Answering RQ7, as the results highlighted, the users consider Emotional Perspective features relevant in the access and search for videos.

Features in Task:	Usefulness		Satisfaction		Ease of Use	
	M	σ	M	σ	M	σ
T25 Search through the EmoMap	4.8	2.2	4.9	1.2	4.8	0.9
T26 Search through emotional filters	4.3	1.1	4.5	1.1	4.9	0.7
T27 Search through the camera	4.6	1.3	4.6	1.4	4.9	0.9
T28 View results in the cover-flow	4.3	0.6	4.0	0.6	4.9	0.3
T29 View results in the map	4.6	0.4	4.6	0.5	4.8	0.3
T30 Information in each graph	4.6	0.3	4.5	0.4	4.9	0.4
T31 Choose a recommended video	4.4	0.4	4.4	0.4	4.9	0.3

Table 5.13: USE evaluation of Emotional features (scale 1-5)

Features in Task:	SAM						PR			
	Pleasure		Arousal		Dominance		Presence		Realism	
	M	σ	M	σ	M	σ	M	σ	M	σ
T25 Search through the EmoMap	8.7	0.5	7.8	0.7	8.3	0.5	8.1	0.8	7.8	0.8
T26 Search through emotional filters	7.6	1.9	6.3	1.4	8.8	2.1	7.7	0.9	7.4	0.7
T27 Search through the camera	8.7	1.7	8.9	1.5	7.4	2.2	8.3	1.0	8.5	0.9
T28 View results in the cover-flow	8.2	1.1	7.4	0.8	8.2	0.9	8.1	1.0	7.9	0.7
T29 View results in the map	8.5	1.0	8.0	1.1	8.4	1.0	8.3	1.3	8.0	0.9
T30 Information in each chart	7.8	0.8	7.5	0.9	8.3	1.1	6.0	0.8	6.1	1.0
T31 Choose a recommended video	5.9	0.7	7.7	0.9	7.7	0.9	6.1	1.3	6.2	1.1

Table 5.14: SAM and PR evaluation of Emotional features (scale: 1-9)

5.2.6 Interaction with TV's & Wider Screens Features

While evaluating the interaction with TV & wider screens features, users were asked to connect the mobile application to a TV set, select a 360° video and navigate it through both the “steering wheel” and the minimap displayed on the tablet (T32). Answering RQ8, users consider that interaction with TVs & wider screens can highly

contribute to a more immersive video environment, as it allows sharing the experience with those around the user, and bigger screens are able to add to the viewing experience. The minimap was especially appreciated, as it provided the user with a reference to the full 360° angle. Most users preferred the navigation through the minimap, stating it was a very intuitive and useful interface. Regarding the “steering wheel” feature, in spite of giving positive comments about its usefulness by stating that it provided a realistic and efficient method to pan around the video without have to look to the mobile device (and thus keeping their attention in the video displayed on the TV), users pointed out that its precision should be carefully adjusted, so that the application disregards false detections (it must successfully distinguish the angle changes that correspond to a turn by the user and the ones which simply represent ordinary hand movement). This happened even when they were told that they could adjust the precision level to their own preferences (as the sensibility of this filter can be adjusted in the user preferences menu). This suggests that the default configuration needs to be less sensible to movement, and that this precision level could be calculated according to a more sophisticated method, possibly through a calibration in the user preferences. After that, users were asked to change direction to a crossing route through a hyperlink using the minimap (T33), and to change to different routes by touching and dragging the marker over trajectories (T34). All users completed the tasks without problems, and they highlighted the marker drag feature on the map, as an interesting “rewind/fast forward” method, where users can advance in the geographical dimension instead of the usual time dimension.

Although this module was evaluated in evaluations one and two, task 32 was the only task conducted during evaluation one, as the features related to tasks 33 and 34 were still in development at the time of evaluation one. The reason for the slight improvements between evaluation one and two may be related to the fact that the sensitivity of the “steering wheel” feature was recalibrated in order to behave in a more fluid manner.

Features in Task:	Usefulness		Satisfaction		Ease of Use	
	M	σ	M	σ	M	σ
T32 Steering wheel and minimap	<i>4.8</i>	<i>0.4</i>	<i>4.8</i>	<i>0.4</i>	<i>4.6</i>	<i>0.5</i>
	4.9	0.5	4.9	0.3	4.6	0.7
T33 Hyperlink through the minimap	-	-	-	-	-	-
	4.7	0.9	4.6	0.6	4.6	0.7
T34 Dragging the marker over trajectories	-	-	-	-	-	-
	4.7	0.5	4.6	0.4	4.6	0.7

Table 5.15: USE evaluation of Interaction with TV's & Wider Screens features (scale 1-5), (evaluation one in italics, evaluation two in bold)

When comparing this interface with the standalone mobile application, users highlighted the possibility that this interface creates to share the video viewing experience with the people around the user. On the other side, users indicated they would use the standalone mobile application more often, as it offers some extra features, such as the wind accessory, which create a more fun and realistic experience.

Features in Task:	SAM						PR			
	Pleasure		Arousal		Dominance		Presence		Realism	
	M	σ	M	σ	M	σ	M	σ	M	σ
T32 Steering wheel and minimap	<i>8.0</i>	<i>0.8</i>	<i>8.7</i>	<i>0.6</i>	<i>7.9</i>	<i>0.5</i>	<i>8.8</i>	<i>0.6</i>	<i>8.7</i>	<i>0.5</i>
	8.2	0.7	8.8	0.8	8.3	0.7	8.9	0.5	8.8	0.4
T33 Hyperlink through the minimap	-	-	-	-	-	-	-	-	-	-
	<i>7.7</i>	<i>0.8</i>	<i>7.7</i>	<i>0.9</i>	<i>7.9</i>	<i>0.7</i>	<i>7.6</i>	<i>0.9</i>	<i>7.5</i>	<i>1.1</i>
T34 Dragging the marker over trajectories	-	-	-	-	-	-	-	-	-	-
	<i>8.0</i>	<i>1.0</i>	<i>7.9</i>	<i>0.9</i>	<i>8.1</i>	<i>0.9</i>	<i>8.0</i>	<i>1.1</i>	<i>8.1</i>	<i>1.2</i>

Table 5.16: SAM and PR evaluation of Interaction with TV's & Wider Screens features (scale: 1-9), (evaluation one in italics, evaluation two in bold)

5.2.7 Evaluating the Emotional Impact of Windy Sight Surfers

The Emotion Recognizer was used itself as a tool to collect data during the entire user evaluation process. In each task, the dominant emotions were logged. Since the objective was to analyse the users' mood while interacting with the application, in the tasks that included video viewing, emotions were not logged while the video was playing. As figure 5.17 shows, the user's dominant mood during the several stages of the evaluation was always related to positive emotions. In fact, negative emotions were not detected by the Emotion Recognizer, being the exceptions tasks T10 and T14, where the Emotion Recognizer detected 'Disgust'.

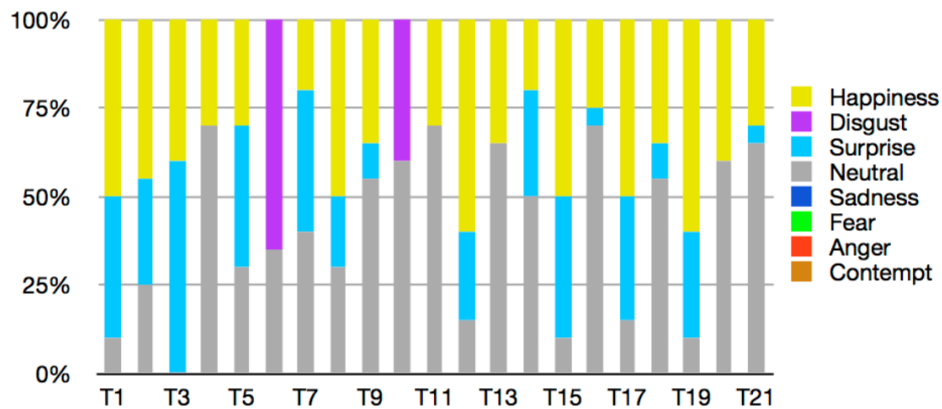


Figure 5.2: Recognized Emotions during user Evaluation. (only Happiness, Disgust, Surprise and Neutral were detected)

Task T10 corresponded to the experience of using a car's engine sound for the Cyclic Doppler Effect, which was not appreciated by the users; T14 corresponded to the experience of using the Cyclic Doppler Effect in a video with very little movement, which proven to be a situation where this feature becomes obtrusive rather than the opposite. This reinforces the conclusion based on user opinions.

5.2.8 Global Presence and Immersion Evaluation

Both the Immersive Tendencies Questionnaire and the Presence Questionnaire were applied in this evaluation in a seven-point scale format.

Tendency	M	σ
Maintain focus on current activities	<i>4.2</i> 4.2	<i>1.5</i> 1.3
Become involved activities	<i>4.3</i> 4.3	<i>1.6</i> 1.7
View videos	<i>5.1</i> 5.0	<i>1.4</i> 1.2

Table 5.17: Immersive Tendencies Questionnaire (scale 1-7), (evaluation one in italics, evaluation two in bold)

Major factor category	M	σ
Control factors	<i>6.1</i> 6.1	<i>0.9</i> 0.8
Sensory factors	<i>6.4</i> 6.7	<i>0.7</i> 0.7
Distraction factors	<i>5.3</i> 5.3	<i>0.9</i> 1.0
Realism factors	<i>5.6</i> 6.3	<i>0.8</i> 0.8
Involvement/Control	<i>6.2</i> 6.2	<i>0.9</i> 1.1
Natural	<i>6.2</i> 6.2	<i>0.8</i> 0.9
Interface quality	<i>6.1</i> 6.2	<i>0.9</i> 0.8

Table 5.18: Presence Questionnaire (scale 1-7), (evaluation one in italics, evaluation two in bold)

Regarding the Immersive Tendencies Questionnaire, each of its questions is directly related with one specific tendency from those outlined in table 5.17. Regarding the Presence Questionnaire, each of its questions is directly related with at least one of the major factor categories outlined in table 5.18.

The Immersive Tendencies Questionnaire revealed a slightly above average score, whereas the Presence Questionnaire showed a high degree of self-reported presence in the application (Tables 5.17 and 5.18). As Presence is a human reaction to immersion, the PQ score reveals the global immersiveness of the tested features. Between evaluations one and two, the major differences in the results are related to the Sensory factors and Realism factors, which indicate improvements in this fields and thus reflect the benefit of the features introduced in evaluation two. More importantly, the improvement from the ITQ to the PQ reveals that the Windy Sight Surfers system surpassed user's usual immersive expectations of a system of this category.

5.2.9 Final Overview

The conducted user evaluation allowed to draw several preliminary conclusions relatively to the several components of the Windy Sight Surfers application, as well as regarding the global immersive capabilities of the designed environment. Besides addressing all the designed features, evaluation two enabled to analyse improvements related to certain features that were refined between evaluation one and two, and allowed to validate the results from evaluation one. It was not possible to gather a large amount of people for the user evaluation, and some features of this application are required to rely on a highly populated 360° video database, which is yet to be achieved. Furthermore, the emotional features are designed to improve their accuracy accordingly to the increasing amount of videos viewed by users, which would require a longer testing period, and was also not possible. Therefore, conclusions must be contextualized, and not declared as absolute truths. Nonetheless, the results were very positive and encouraging, being that the vast majority of the conclusions comprised positive responses to the proposals of the designed application.

Chapter 6

Conclusions and Future Work

This chapter presents the final thoughts on the work described throughout this thesis. The key decisions and results are analysed, followed by a summarization of the major contributions of this work. Afterwards, some directions for future work are discussed.

6.1 Conclusions

Having the motivation and research objectives being identified, the state of the art of the areas more closely related to the work described in this thesis was studied. In order to achieve the referred objectives, a number of solutions was proposed, implemented and subject to user evaluation. The outcome of this process resulted in the Windy Sight Surfers application.

The capture of 360° video was enhanced through the collection of associated metadata, which enabled to reproduce the captured 360° videos in a more immersive environment, as additional information could be given to the user about the real scenario portrayed in the captured video. Also, this information was used in the design of several video search processes. Using the information obtained in the video capture process, a spatial dimension was added to the search process, where users can search for videos in certain locations.

For the reproduction of 360° videos, the Windy Sight Surfers application introduced different components that strive to enhance the immersiveness of the video viewing experience. Namely, the 360° videos were mapped onto a transitional canvas that was in turn rendered around a cylinder, to represent the 360° view and allow the feeling of being “partially” surrounded by the video. Being the Windy Sight Surfers an application contextualized in mobile environments, the application enabled to change the video viewing angle either by moving the mobile device around, as if it was a window to the video, or by using the screen as a panning interface.

In order to create a more realistic perception of speed and movement while viewing 360° videos, a Wind Accessory was designed and developed, consisting of a device that blows wind to the viewer. Once the video's metadata file contains information about speed and wind, the Wind Accessory operates its fans in real time according to messages received from the Windy Sight Surfers application, which contain values that are based on the video's metadata. As this application focuses on 360° video, the wind produced by the wind accessory also takes the orientation parameter into account, so that the wind speed is higher when the user is viewing the video angle that opposes to the captured wind direction.

Further exploration was conducted regarding sound and its applicability with respect to improving user orientation when viewing 360° videos and improving the perception of speed and movement. In this context, the video's sound was mapped onto a 3D Sound Space. The motivation behind this decision is that, while 360° videos are reproduced as if the tablet was a window to the video surrounding the user (meaning that a specific video angle is being viewed at a time), when a video is recorded the sound is recorded accordingly to the orientation of the camera, which results in the static reproduction of the video's sound component (the sound is always the same regardless of the angle being viewed). Therefore, this approach was presented as a means to allow the user to easily identify the orientation of the video through sound. Regarding the improvement of the perception of speed and movement through sound, a second sound layer was added to the video, which cyclically reproduces the Doppler Effect in a controlled manner (also in a 3D space). Different sounds and volumes (being the volume of the Doppler Effect sound layer established in comparison to the video's sound layer) were experimented in the Doppler Effect simulation. Also, situations of videos with different movement degrees were analysed in order to find which are the cases where the Doppler Effect simulation becomes beneficial (our findings indicate there is a minimum amount of movement required for the Doppler Effect to become beneficial).

Being that immersion values can be increased through context awareness, an overlay is always present above the videos while they are being reproduced. This overlay enriches the video with additional features that provide information from the video's metadata and are divided in two categories: Permanent Information Features, which relate to items of information that are permanently present on the screen throughout the video reproduction; and Momentary Information Features, which relate to items of information that appear momentarily on the screen and are related to specific portions of the video.

Focusing on a more engaging experience to the user, an emotion recognizer system was incorporated into the application. Based on a Facial Expression Recognition Framework, this system was used in the context of the Windy Sight Surfers application as a means to recognize the user's emotions when visualizing video and using the application. Capturing the user's facial expression while viewing videos, the Emotion Recognizer processed the captured images, associating an expression with each one of them. This emotional information has three applications in the Windy Sight Surfers environment, being used in features related to Video Emotional Cataloguing and Access (on the map and through search), in features that may influence the control flow of the environment, and in the user evaluation of the environment itself, based on the emotional impact.

Striving to increase the immersiveness of the environment, a module was added to the system, providing the mobile application with the capability to interact with wider screens, such as TVs, and take advantage of their screen size. In this context, videos are reproduced in full screen on the TV and the mobile application acts as a second screen, being the only responsible for the control of the multi-screen application. Furthermore, the mobile application is used as an extension of the video content displayed on the TV, providing additional information and content.

The user evaluation showed that Windy Sight Surfers increases the sense of presence and immersion, and that users appreciated the designed features, finding them useful, satisfactory and easy to use. Users especially showed great interest in the wind accessory, noting its effectiveness regarding the improvement of the realism of the environment. Also, users especially liked the video navigation by moving the tablet around, which they reported to be a more natural approach when compared to the more familiar drag interface. The work comprised in this thesis was accepted in the scientific community, as it resulted in the publication of a full paper in the international reference conference on interactive TV and video (url-EuroITV), and a full paper in the Immersive Media Experiences workshop at ACM MM'13, the premier international conference on multimedia.

6.2 Future Work

Future developments include populating the system with a wide variety of 360° videos. This will enable to build up the Windy Sight Surfers application and do further testing with users, especially relating to the emotionally based video recommendation system, as its effectiveness study requires a large amount of different videos and also a large group of users. This will provide the means to see in a more precise way what types of techniques are more useful.

Once the application reaches a publishable stage, social networks integration would be a very interesting approach, not only because it would be a way to publicise the application, thus obtaining a larger number of users and, consequently, a much larger quantity of information regarding the application's usage. Therefore, this would be a plausible way to test and refine the recommendation system.

There is also interest in refining and extending the current solutions by exploring further more immersive settings, like wide projection screens or the CAVE, trying to understand what the role of mobile devices in these highly immersive environments might be.

Regarding the system implementation, it is important to adapt the emotion recognizer framework (also developed in the scope of the ImTV research project) to mobile platforms, such as Android. This must be done in order to dismiss the requirement of the application to be connected to a computer and webcam in order to recognize the user's facial expressions, and take the application to a publishable stage.

An interesting approach might be to extend the Cyclic Doppler Effect feature so that it takes into account the orientation of the video viewing angle. More specifically, considering figure 3.18, with this extension the Cyclic Doppler Effect would still approach, pass, and recede the users' head, although the effect would not execute explicitly from the front to the back, but the orientation of it would rather be relative to the orientation of the video viewing angle. The main objective would be to investigate whether this functionality would allow the user to easily identify the orientation of the movement in the video while it is being reproduced, and if this extension would not increase the complexity of the reproduced effect to such an extent that it becomes obtrusive rather than beneficial.

Still discussing the system implementation, it is also relevant to try to find an alternative to the Web Audio API used through the connexion to a computer, thus eliminating the requirement of the application to be connected to a computer and webcam in order to reproduce the sounds in a 3D space. The ideal scenario would be the implementation of an Android Java library, which would simply be included in the Windy Sight Surfers application. But, as in present times that option does not seem viable, the second alternative would be to implement a service application in the Android's low level Native Development Kit, which would run under the Windy Sight Surfers Android application. Another option is to keep the implementation of the 3D audio component based on the Web Audio API, but use it in a WebView component. Despite being the most desirable alternative, this alternative still requires too much computing power when considering mobile devices. Nevertheless, in the near future, this can be considered the most attractive approach, as it enables to reuse the implemented code, thus making the transition process relatively easy.

Bibliography

Alliance for Telecommunications Industry Solutions. ATIS IPTV Exploratory Group Report and Recommendation to the TOPS Council. July 2005.

Almeida, P., Abreu, J., Pinho, A., Costa, D., Engaging Viewers through Social TV Games. In EuroiTV '12 Proceedings of the tenth international interactive conference on Interactive Television. ACM (2012).

Álvares, C., 2012. Vídeos Interativos e Imersivos no Sight Surfers, MSc thesis, Universidade de Lisboa, September.

Aroyo, L., Nixon, L., Miller, L., NoTube: the television experience enhanced by online social and semantic data. In 1st International Conference on Consumer Electronics (IC-CE 2011, Berlin, Germany, September).

Baker, H., Chang, N., Paruchuri, A., Capture and Display for Live Immersive 3D Entertainment. In MM '11 Proceedings of the 19th ACM international conference on multimedia. ACM(2011).

Bernhaupt, R., Boutonnet, M., Gatellier, B., Gimenez, Y., Pouchepanadin, C., Souiba, L., A Set of Recommendations for the Control of IPTV-Systems via Smart Phones based on the Understanding of Users Practices and Needs. In EuroiTV '12 Proceedings of the tenth international interactive conference on Interactive Television. ACM (2012).

Bleumers, L., Broeck, W., Lievens, B., Pierson, J., Seeing the Bigger Picture: A User Perspective on 360° TV. In EuroiTV '12 Proceedings of the tenth international interactive conference on Interactive Television. ACM (2012).

Brave, S., Nass, C. Emotion in human-computer interaction, The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications, Lawrence Erlbaum Associates, Inc., Mahwah, NJ (2002).

Brondmo, H., Davenport, G., Creating and Viewing the Elastic Charles – a Hypermedia Journal. In Hypertext2, York, England (1989).

Brown, E., Cairns, P., A Grounded Investigation of Game Immersion. In Proceedings of CHI EA '04 Extended Abstracts on Human Factors on Computing Systems. ACM (2004).

Bulterman, D., Rossum, G., Liere, R., A Structure for Transportable, Dynamic Multimedia Documents. In Proceedings of the Summer 1991 USENIX Conference (1991).

Cardin, S., Thalmann, D., and Vexo, F. 2007, Head Mounted Wind. In Proceedings of Computer Animation and Social Agents. Hasselt, Belgium, June 11-13. pp.101-108.

Chambel, T., Langlois, T., Martins, P. "MovieClouds: Content-Based Overviews and Exploratory Browsing of Movies". In Proceedings of Academic MindTrek' 2011, in Cooperation with ACM SIGCHI & SIGMM, pp.133-140, Tampere, Finland, Sep 28-30, 2011.

Dakss, J., Agamanolis, S., Chalom, E., Bove Jr, V., Hyperlinked Video. In SMPTE Motion Imaging Journal (1998).

Department of Industry, London (United Kingdom), World System Teletext and Data Broadcasting System Technical Specification. In The British Library (1989).

Dobler, D., and Stampfl, P. (2004) Enhancing Three-dimensional Vision with Three-dimensional Sound. In ACM Siggraph'04 Course Notes.

Douglas, Y., and Hargadon, A., 2000. The Pleasure Principle: Immersion, Engagement, Flow. ACM Hypertext'00. San Antonio, Texas, USA, pp.153-160.

Ekman, P. Are there basic emotions? Psychological Review, 99(3): 550-553 (1992).

Ferman, A., Beek, P., Errico, J., Sezan, M., Multimedia Content Recommendation Engine with Automatic Inference of User Preferences. In Proceedings IEEE International Conference on Image Processing (Barcelona, Spain, 14-17 September). IEEE (2003).

Girgensohn, A., Shipman, F., Wilcox, L., Hyper-Hitchcock: Authoring Interactive Videos and Generating Interactive Summaries. In MM '03 Proceedings of the eleventh ACM international conference on multimedia. ACM (2003).

Halasz, F., Schwartz, M. The Dexter hypertext reference model. Communications of the ACM, Volume 37 Issue 2, Feb. 1994. pp.30-39.

Hardman, L., Modelling and Authoring Hypermedia Documents. Phd Thesis, University of Amsterdam (1998). Chapter 3.

Heilig, M. Sensorama Simulator. U.S. Patent #3,050,870. August 1962.

Hirata, K., Takano, H., Hara, Y., Miyabi: A Hypertext Database with Media-Based Navigation (Video). ACM Hypertext. pp.233-234 (1993).

Huhtamo, E. Encapsulated Bodies in Motion. In: Penny, Simon (ed.). Critical Issues in Electronic Media. New York: State University of New York (1995). pp.159-186.

Jonietz, E., Making TV Social, Virtually. In MIT Technology Review (2010).

Kleinginna, P.R., & Kleinginna, A.M. A categorized list of emotion definitions with suggestions for a consensual definition. Motivation and Emotion. Volume 5, Issue 4 pp.345-379, December 1981.

Kojima, Y., Hashimoto, Y., Kajimoto, H. 2009. A novel wearable device to present localized sensation of wind. In Proc. of ACE'09. Rome, Italy, January 12-14, pp.61-65.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), Technology in mental health care delivery systems (pp. 119-137). Norwood, NJ: Ablex.

Lehmann, A., Geiger, C., Wöldecke, B. and Stöcklein, J. 2009, Poster: Design and Evaluation of 3D Content with Wind Output. In Proceedings of 3DUI'09. Lafayette, Louisiana, USA, March, pp.14-15.

Lochrie, M., Coulton, P., Sharing the viewing experience through Second Screens. In EuroiTV '12 Proceedings of the tenth international interactive conference on Interactive Television. ACM (2012).

Lund, A. M. "Measuring usability with the USE questionnaire. Usability and User Experience", 8(2). 8. 2001.

Manuel, D., Moore, D., and Charissis, V. (2012) An Investigation into Immersion in Games Through Motion Control and Stereo Audio Reproduction. In Proc. of AM '12: 7th Audio Mostly Conference, Corfu, Greece, Sep 26-28. pp.124-129.

Martins, F., Peleja, F., Magalhães, J., SentiTVchat: Sensing the Mood of Social-TV Viewers. In EuroiTV '12 Proceedings of the tenth international interactive conference on Interactive Television. ACM (2012).

McMahan, A., Immersion, Engagement, and Presence – A Method for Analyzing 3-D Video Games. Routledge Chapman & Hall Verlag (2003).

Mendes, M. 2010. RTiVISS | Real-Time Video Interactive Systems for Sustainability. In Proceedings of Artech'10. Guimarães, Portugal, April 22-23, pp.29-38.

Merer, A., Aramaki, M., Ystad, S., and Kronland-Martinet, R. 2013. Perceptual Characterization of Motion Evoked by Sounds for Synthesis Control Purposes. In Transactions on Applied Perception (TAP), 10(1). pp.1-23.

Moon, T., and Kim, G. J. 2004, Design and evaluation of a wind display for virtual reality. In Proceedings of the ACM Symposium on Virtual Reality Software and Technology. Hong Kong, China, November 10-12. pp.122-128.

Mourão, A., Borges, P., Correia, N., Magalhães, J., Facial Expression Recognition by Sparse Reconstruction with Robust Features. In Image Analysis and Recognition (ICIAR'2013), LNCS, vol 7950, pp.107- 115, 2013.

Murray, J., Hamlet on the Holodeck. The MIT Press (1997).

Nazemi, M., and Gromala, D. 2012. Sound Design: A Procedural Communication Model for VE. In Proc. of AM '12: 7th Audio Mostly Conference, Corfu, Greece. pp.16-23.

Nelson, T., Branching presentational systems: Hypermedia. In Dream Machines (1974). pp.44-45.

Neng, L., 2010. 360° Hypervideo, MSc thesis, Universidade de Lisboa, September.

Neng, L., and Chambel, T., 2010. Get Around 360° Hypervideo. In Proc.of MindTrek (2010). Tampere, Finland, October 6-8, pp.119-122.

Neng, L., Chambel, T., Chhanganlal, M., Towards Immersive Interactive Video Through 360° Hypervideo. In ACE'2011 (Lisbon, Portugal, 8-11 November). ACM (2011).

Noronha, G., 2012. Sight Surfers: Partilha e Geonavegação em Vídeos 360°, MSc thesis, Universidade de Lisboa, September.

Noronha, G., Álvares, C., and Chambel, T., 2012. Sight Surfers: 360° Videos and Maps Navigation. In Proc.of GeoMM'12 ACM Multimedia'12. Nara, Japan, October 29-November 2, pp.19-22.

Oliveira, E., Martins, P., Chambel, T., "Accessing Movies Based on Emotional Impact", Special Issue on "Social Recommendation and Delivery Systems for Video and TV Content", ACM/Springer Multimedia Systems Journal, June 2013.

Oliveira, E., Martins, P., Chambel, T., "iFelt: Accessing Movies Through Our Emotions". In Proceedings of EuroITV'2011: "9th International Conference on Interactive TV and Video: Ubiquitous TV", in cooperation with ACM SIGWEB, SIGMM & SIGCHI, pp.105-114, Lisbon, Portugal, June 29-July 1, 2011.

Plutchik, R. Emotion: A Psychoevolutionary Synthesis. Harpercollins College Div, 1980.

Prata, A., Chambel, T., Mobility in a Personalized and Flexible Video Based Trans-media Environment. (2011) In Ubicomm'2011 (Lisbon, Portugal, 20-25 November).

Russell, J. A circumflex model of affect. *Journal of Personality and Social Psychology*, Issue 39. pp.1161–1178, 1980.

Sawhney, N., Balcom, D., Smith, I., HyperCafe: Narrative and Aesthetic Proprieties of hypervideo. In *Hypertext 1996*. ACM(1996).

Sekar, V., Dobrian, F., Awan, A., Joseph, D., Ganjam, A., Zhan, J., Stoica, I., Zhang, H., Understanding the Impact of Video Quality on User Engagement. In *SIGCOMM '11 Proceedings of the ACM SIGCOMM 2011 conference*. ACM (2011).

Slater, M., Lotto, B., Arnold, M.M., Sanchez-Vives, M.M. (2009) How we experience immersive virtual environments: the concept of presence and its measurement, *Anuario de Psicología*, 40(2), Fac. Psicologia, Univ. Barcelona, pp.193-210.

Strover, S., and Moner, W., “Immersive television and the on-demand audience”. Presented at the International Communication Association Conference, May 2012, Phoenix, USA.

Suk, H. (2006). Color and Emotion - a study on the affective judgment across media and in relation to visual stimuli. University of Mannheim: Dissertation.

Tseklevs, E., Cruickshank, L., Hill, A., Kondo, K., Whitham, R., Interacting with Digital Media at Home via a Second Screen. In *Proceedings of the ISMW '07 of the Ninth IEEE International Symposium on Multimedia Workshops*. IEEE (2007).

Visch, T., Tan, S., and Molenaar, D., 2010. The emotional and cognitive effect of immersion in film viewing. *Cognition & Emotion*, 24: 8, pp.1439-1445.

Wei, D., Scheuerer, J., Wenleder, M., Erk, A., Glbahar, M., Linnhoff-Popien, C., A User Profile-based Personalization System for Digital Multimedia Content. In *DIMEA'08 (Athens, Greece, 10-12 September)*. ACM (2008). pp.281-288

Witmer, B., Singer, M.J. (1998) Measuring Presence in Virtual, Environments: A Presence Questionnaire. *Presence*, 7(3), Jun. (1998), MIT, 225–240.

Internet References

- (url-ACM) ACM 2013 Conference: <http://acmmm13.org/>
- (url-AkaiSynthStation) <http://www.akaipro.com/synthstation>
- (url-AndroidAOA) Android AOA Protocol:
<http://source.android.com/accessories/protocol.html>
- (url-AndroidNDK) Android NDK:
<http://developer.android.com/tools/sdk/ndk/index.html>
- (url-APDC) Congresso das Comunicações da APDC: <http://congresso12.apdc.pt/>
- (url-AppleTV) Apple TV: <http://www.apple.com/appletv/>
- (url-Arduino) Arduino: <http://www.arduino.cc/>
- (url-ArduinoMegaADK) Arduino Mega ADK:
<http://arduino.cc/en/Main/ArduinoBoardADK>
- (url-ATC9K) ATC9K:
<http://uk.oregonscientific.com/cat-Outdoor-sub-Action-Cam-prod-ATC9K-HD-Action-Camera.html>
- (url-ATIS) Alliance for Telecommunications Industry Solutions:
<http://www.atis.org/>
- (url-BeaufortScale) Beaufort Scale:
http://en.wikipedia.org/wiki/Beaufort_scale#Modern_scale
- (url-BuzMuzik) BuzMuzik:
<http://www.cscmediagroup.com/showreels/367/BuzMuzik>
- (url-Captcha) Captcha: <http://www.captcha.net/>
- (url-CiscoGlobalIPTraffic) Cisco Global IP Traffic Forecast:
http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html
- (url-CodeIgniter) CodeIgniter: <http://ellislab.com/codeigniter>

(url-DirectSound) Direct Sound:

[http://msdn.microsoft.com/en-us/library/windows/desktop/ee416960\(v=vs.85\).aspx](http://msdn.microsoft.com/en-us/library/windows/desktop/ee416960(v=vs.85).aspx)

(url-DirectX) Microsoft DirectX SDK:

<http://windows.microsoft.com/en-us/windows7/products/features/directx-11>

(url-EuroITV) EuroITV: <http://www.euro-itv.org/>

(url-FacebookLogin) Facebook Login:

<https://developers.facebook.com/docs/facebook-login/>

(url-FCULDiaAberto) FCUL Dia Aberto:

<http://www.fc.ul.pt/pt/pagina/1932/dia-aberto>

(url-FFmpeg) FFmpeg: <http://ffmpeg.org/>

(url-GenC) Generation-C:

www.nielsen.com/us/en/newswire/2012/introducing-generation-c.html

(url-GoogleGenC) Google on Generation-C:

<http://adwordsagency.blogspot.pt/2013/03/how-does-gen-c-watch-youtube-on-all.html>

(url-GoogleMaps) Google Maps: <https://maps.google.com/>

(url-GoogleMapsAndroidAPI) Google Maps Android API:

<https://developers.google.com/maps/documentation/android/>

(url-GoogleMapsAndroidAPIZoomLevels) Google Maps Android API Changing Zoom Levels:

https://developers.google.com/maps/documentation/android/views#changing_zoom_level

(url-GoogleMapsAPI) Google Maps API: link!

<https://developers.google.com/maps/>

(url-GSON) Google's GSON Java Library: <https://code.google.com/p/google-gson/>

(url-HCIM) HCIM: <http://hcm.di.fc.ul.pt/>

(url-HotVideo) HotVideo:

<http://www.research.ibm.com/topics/popups/innovate/multimedia/html/hotvideo.html>

(url-HyperSoap) HyperSoap: <http://www.media.mit.edu/hypersoap/>

(url-iCloud) Apple iCloud: <http://www.apple.com/icloud/>

(url-IDC) International Data Corporation:
<http://www.idc.com/about/viewpressrelease.jsp?containerId=prUS22689111>

(url-IMAX) IMAX: <http://www.imax.com/>

(url-ImmersiveME) Immersive Media Experiences ACM Workshop:
<http://immersiveme2013.di.fc.ul.pt/>

(url-ImTV) ImTV: <http://imtv.me/>

(url-ImTVWorkshop) Second ImTV Workshop:
<http://imtv.me/imtv-2nd-workshop-meeting-245/>

(url-Iosono) Iosono: link: <http://www.iosono-sound.com/>

(url-JavaDOM) Java's DOM Parser:
<http://docs.oracle.com/javase/6/docs/api/javax/xml/parsers/DocumentBuilder.html>

(url-JavaNativeInterface) Java Native Interface:
<http://docs.oracle.com/javase/7/docs/technotes/guides/jni/spec/intro.html#wp9502>

(url-Java3DSound) Java 3D Sound API:
<http://download.java.net/media/java3d/javadoc/1.4.0/javax/media/j3d/doc-files/intro.html>

(url-JSON) JSON: <http://www.json.org/>

(url-JWebSocket) JWebSocket: <http://jwebsocket.org/>

(url-KingstonVOD) Kingston VOD System:
http://www.broadcastpapers.com/whitepapers/casestudy_kingston.pdf?CFID=24446886&CFTOKEN=d3d2ef8bbf91d3e-42CE595D-E03F-82D2-FC8CBDBBCC72DC53

(url-LaSIGE) LaSIGE: <http://lasige.di.fc.ul.pt/>

(url-Line6StageScape) Line 6 StageScape: <http://line6.com/stagescape-m20d/>

(url-LogitechWebcam) Logitech Webcam:
<http://www.logitech.com/en-us/product/hd-webcam-c310>

(url-Mappiness) Mappiness: <http://www.mappiness.org.uk/>

(url-Melies) Georges Méliès: <http://www.melies.eu/English.html>

(url-MHP) Multimedia Home Platform: <http://www.mhp.org/>

(url-Miso) Miso: <http://www.gomiso.com>

(url-Moog) Moog Little Phatty: <http://www.moogmusic.com/products/phattys/little-phatty-stage-ii>

(url-MySQL) MySQL: <http://www.mysql.com/>

(url-Nginx) Nginx: <http://wiki.nginx.org/Main>

(url-Nike+) Nike+: http://nikeplus.nike.com/plus/products/gps_app/

(url-NoTube) NoTube: <http://notube.tv/>

(url-OpenAL) OpenAL: <http://connect.creativelabs.com/openal/default.aspx>

(url-OpenGL) OpenGL: <http://connect.creativelabs.com/openal/default.aspx>

(url-OpenWeatherMap) OpenWeatherMap: <http://openweathermap.org/>

(url-OpenWeatherMapWeatherCodes) OpenWeatherMap weather condition codes: http://openweathermap.org/wiki/API/Weather_Condition_Codes

(url-PaulNipkow) Paul Nipkow: <http://inventors.about.com/od/germaninventors/a/Nipkow.htm>

(url-PlayAlong) PlayAlong: <http://www.visiware.com/?p=675>

(url-SAM) SAM: http://irtel.uni-mannheim.de/pxlab/demos/index_SAM.html

(url-SMIL) SMIL: <http://www.w3.org/TR/SMIL/>

(url-SparkFunDualMotorDriver) SparkFun Dual Motor Driver: <https://www.sparkfun.com/products/9457>

(url-SurfaceView) Android SurfaceView class: <http://developer.android.com/reference/android/view/SurfaceView.html>

(url-TodayShow) Today Show: <http://today.msnbc.msn.com/>

(url-TVCaboDiTV) TV Cabo and Microsoft DiTV Partnership: <http://www.microsoft.com/en-us/news/press/2001/may01/05-17tvcabopr.aspx>

(url-TVcheck) TVcheck: <http://tvcheck.com/uk/>

(url-TwitterLogin) Facebook Login: <https://dev.twitter.com/docs/auth/sign-twitter>

(url-VideoClix) VideoClix: <http://www.videoclix.tv/>

(url-VideoView) Android VideoView class: <http://developer.android.com/reference/android/widget/VideoView.html>

(url-WebAudioAPI) Web Audio API: <https://dvcs.w3.org/hg/audio/raw-file/tip/webaudio/specification.html>

(url-WebGL) WebGL: <http://www.khronos.org/webgl/>

(url-WeFeelFine) We Feel Fine: <http://www.wefeelfine.org/>

(url-WiiU) Nintendo Wii U: <http://www.nintendo.com/wiiu>

(url-WiO) WiO: <https://wioffer.com/wio/about/>

(url-W3C) W3C: <http://www.w3.org/>

(url-XboxSmartGlass) Xbox SmartGlass:

<http://www.xbox.com/en-US/live/smartglass>

(url-YouTube) YouTube: <http://www.youtube.com/>

(url-YouTubeAnnotations) YouTube Annotations:

http://www.youtube.com/t/annotations_about

(url-YouTubeReport) YoutubeReport:

<http://youtube-global.blogspot.pt/2013/03/onebillionstrong.html>

(url-YouTubeStatistics) YouTube Statistics:

<http://www.youtube.com/yt/press/statistics.html>

Annex A: Video's Metadata File Example

```
<?xml version="1.0" encoding="UTF-8"?>

<root>

  <wind>

    <name>Lisbon</name>

    <windspeed>0.1</windspeed>

    <deg>179</deg>

    <weathercode>800</weathercode>

  </wind>

  <coordinates>

    <coordinate>

      <lat>38.713323</lat>

      <lon>-9.139123</lon>

      <speed>5.2</speed>

      <orientation>129.20808</orientation>

      <timestamp>1365513122</timestamp>

      <date>2013-04-09 13:12:02.83</date>

    </coordinate>

    <coordinate>

      <lat>38.713299</lat>

      <lon>-9.139111</lon>

      <speed>5.6</speed>

      <orientation>129.20808</orientation>

      <timestamp>1365513123</timestamp>

      <date>2013-04-09 13:12:3.41</date>

    </coordinate>

  </coordinates>

</root>
```


Annex B: User Evaluation One Script

Task	Success?	Usability (U) Usefulness (S) Satisfaction (E) Ease of Use	Self-Assessment-Manikin (P) Pleasure (A) Arousal (D) Dominance	(P) Presence (R) Realism	Obs. / Comments
Perceptual Sensing Features					
1 – Move around a 360° video by moving the tablet around and by using the drag.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
2 – View a 360° video with the Wind Accessory, and ask the user to try to identify the wind direction by moving the tablet around.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
Context Awareness Features					
3 – View a video with the overlay, albeit with only the permanent information activated, and identify each category of information provided.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
4 – View a predefined set of three videos about a specific theme with the overlay that has both the permanent and momentary information activated, and identify the permanent category related to each example of momentary information (all the momentary information available was always presented).	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
5 – Similar to 4, except that the momentary information presented took into account the user preferences.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
6 – While watching a 360° video, find and follow a link to a crossing trajectory.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
Perceptual Sensing & Context Awareness features in Conjunction					
7 – Watch and interact with a video with all of the created features (the perceptual sensing and the context awareness features) activated.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
8 – Watch and interact with a video with the perceptual sensing features disabled.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
9 – Watch and interact with a video with the context awareness features disabled.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
10 – Watch and interact with a video with all the features disabled (just bare video).	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
Interaction with TV's & Wider Screens Features					
11 – Connect the mobile application to a TV set, select a 360° video and navigate it through both the “steering wheel” and the minimap displayed on the tablet.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	

Annex C: User Evaluation Two Script

Task	Success?	Usability (U) Usefulness (S) Satisfaction (E) Ease of Use	Self-Assessment-Manikin (P) Pleasure (A) Arousal (D) Dominance	(P) Presence (R) Realism	Obs. / Comments
Video Search Features					
1 – Conduct a video search of a specified video based on the set of filters provided by the application and on a set of keywords.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
2 – Conduct a video search of a specified video using the “search through map” feature.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
3 – Find and select the specified video through the cover-flow method.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
4 – Find and select the specified video through the map.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
Perceptual Sensing Features					
5 – Move around a 360° video by moving the tablet around (using only this method).	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
6 – Move around a 360° video by using the drag interface (using only this method).	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
7 – View a 360° video with the Wind Accessory, and ask the user to try to identify the wind direction by moving the tablet around.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
8 – View a 360° video with headphones being that the video’s sound was standard stereo sound.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
9 – View the same video with the 3D sound capability.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
10 – View a video with the Cyclic Doppler Effect feature using the sound of a car’s engine.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	
11 – View a video with the Cyclic Doppler Effect feature using the designed low frequency sounds.	Yes No	(U): 1 2 3 4 5 (S): 1 2 3 4 5 (E): 1 2 3 4 5	(P): 1 2 3 4 5 6 7 8 9 (A): 1 2 3 4 5 6 7 8 9 (D): 1 2 3 4 5 6 7 8 9	(P): 1 2 3 4 5 6 7 8 9 (R): 1 2 3 4 5 6 7 8 9	

Task	Success?	Usability (U) Usefulness (S) Satisfaction (E) Ease of Use	Self-Assessment-Manikin (P) Pleasure (A) Arousal (D) Dominance	(P) Presence (R) Realism	Obs. / Comments
Perceptual Sensing Features					
12 – View a video with the Cyclic Doppler Effect feature with High degree of movement.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
13 – View a video with the Cyclic Doppler Effect feature with medium degree of movement.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
14 – View a video with the Cyclic Doppler Effect feature with little degree of movement.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
15 – View a video without the Cyclic Doppler Effect feature.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
16 – View a video with the custom Cyclic Doppler Effect feature.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
Context Awareness Features					
17 – View a video with the overlay, albeit with only the permanent information activated, and identify each category of information provided.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
18 – View a predefined set of three videos about a specific theme with the overlay that has both the permanent and momentary information activated, and identify the permanent category related to each example of momentary information (all the momentary information available was always presented).	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
19 – Similar to 18 , except that the momentary information presented took into account the user preferences.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
20 – While watching a 360° video, find and follow a link to a crossing trajectory.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	

Task	Success?	Usability (U) Usefulness (S) Satisfaction (E) Ease of Use	Self-Assessment-Manikin (P) Pleasure (A) Arousal (D) Dominance	(P) Presence (R) Realism	Obs. / Comments
Perceptual Sensing & Context Awareness features in Conjunction					
21 – Watch and interact with a video with all of the created features (the perceptual sensing and the context awareness features) activated.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
22 – Watch and interact with a video with the perceptual sensing features disabled.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
23 – Watch and interact with a video with the context awareness features disabled.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
24 – Watch and interact with a video with all the features disabled (just bare video).	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
Emotional Features					
25 – Search a video in which the predominant emotion is anger through the EmoMap.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
26 – Search a video in which the predominant emotion is happiness through the checkboxes filter tool.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
27 – Search a video in which the predominant emotion is happiness by exemplifying the desired emotion to the camera.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
28 – Search for neutral videos and view the results in the cover-flow mode.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
29 – Search for neutral videos and view the results in the map.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
30 – Identify what information was contained in each of the graphs and charts.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
31 – Choose and view a recommended video.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
Interaction with TV's & Wider Screens Features					
32 – Connect the mobile application to a TV set, select a 360° video and navigate it through both the “steering wheel” and the minimap displayed on the tablet.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
33 – Users were asked to change direction to a crossing route through a hyperlink using the minimap.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	
34 – Change to different routes by touching and dragging the marker over trajectories.	Yes No	(U): 12345 (S): 12345 (E): 12345	(P): 123456789 (A): 123456789 (D): 123456789	(P): 123456789 (R): 123456789	

Annex D: Immersive Tendencies Questionnaire

Immersive Tendencies Questionnaire						
Indicate your preferred answer by marking an "X" in the appropriate box of the seven-point scale. Please consider the entire scale when making your responses, as the intermediate levels may apply. For example, if your response is once or twice, the second box from the left should be marked. If your response is many times but not extremely often, then the sixth (or second box from the right) should be marked.						
1. Do you ever get extremely involved in projects that are assigned to you by your boss or your instructor, to the exclusion of other tasks?						
NEVER		OCCASIONALLY			OFTEN	
2. How easily can you switch your attention from the task in which you are currently involved to a new task?						
NEVER		OCCASIONALLY			OFTEN	
3. How frequently do you get emotionally involved (angry, sad, or happy) in the news stories that you read or hear?						
NEVER		OCCASIONALLY			OFTEN	
4. How well do you feel today?						
NEVER		OCCASIONALLY			OFTEN	
5. Do you easily become deeply involved in movies or TV dramas?						
NEVER		OCCASIONALLY			OFTEN	
6. Do you ever become so involved in a television program or book that people have problems getting your attention?						
NEVER		OCCASIONALLY			OFTEN	
7. How mentally alert do you feel at the present time?						
NEVER		OCCASIONALLY			OFTEN	
8. Do you ever become so involved in a movie that you are not aware of things happening around you?						
NEVER		OCCASIONALLY			OFTEN	
9. How frequently do you find yourself closely identifying with the characters in a story line?						
NEVER		OCCASIONALLY			OFTEN	
10. Do you ever become so involved in a video that it is as if you are inside the video rather than lying on the couch while watching the screen?						
NEVER		OCCASIONALLY			OFTEN	
11. On average, how many books do you read for enjoyment in a month?						
NEVER		OCCASIONALLY			OFTEN	

12. How physically fit do you feel today?						
NEVER		OCCASIONALLY			OFTEN	
13. How good are you at blocking out external distractions when you are involved in something?						
NEVER		OCCASIONALLY			OFTEN	
14. When watching sports, do you ever become so involved in the game that you react as if you were one of the players?						
NEVER		OCCASIONALLY			OFTEN	
15. Do you ever become so involved in a daydream that you are not aware of things happening around you?						
NEVER		OCCASIONALLY			OFTEN	
16. Do you ever have dreams that are so real that you feel disoriented when you awake?						
NEVER		OCCASIONALLY			OFTEN	
17. When playing sports, do you become so involved in the game that you lose track of time?						
NEVER		OCCASIONALLY			OFTEN	
18. Are you easily disturbed when working on a task?						
NEVER		OCCASIONALLY			OFTEN	
19. How well do you concentrate on enjoyable activities?						
NEVER		OCCASIONALLY			OFTEN	
20. How often do you view videos?						
NEVER		OCCASIONALLY			OFTEN	
21. How well do you concentrate on disagreeable tasks?						
NEVER		OCCASIONALLY			OFTEN	
22. Have you ever gotten excited during a chase or fight scene on TV or in the movies?						
NEVER		OCCASIONALLY			OFTEN	
23. To what extent have you dwelled on personal problems in the last 48 hours?						
NEVER		OCCASIONALLY			OFTEN	
24. Have you ever gotten scared by something happening on a TV show or in a movie?						
NEVER		OCCASIONALLY			OFTEN	

25. Have you ever remained apprehensive or fearful long after watching a scary movie?

--	--	--	--	--	--	--

NEVER

OCCASIONALLY

OFTEN

26. Do you ever avoid carnival or fairground rides because they are too scary?

--	--	--	--	--	--	--

NEVER

OCCASIONALLY

OFTEN

27. How frequently do you watch TV soap operas or docu-dramas?

--	--	--	--	--	--	--

NEVER

OCCASIONALLY

OFTEN

28. Do you ever become so involved in doing something that you lose a track of time?

--	--	--	--	--	--	--

NEVER

OCCASIONALLY

OFTEN

Annex E: Presence Questionnaire

Presence Questionnaire						
Characterize your experience in the environment, by marking an "X" in the appropriate box of the 7-point scale, in accordance with the question content and descriptive labels. Please consider the entire scale when making your responses, as the intermediate levels may apply. Answer the questions independently in the order that they appear. Do not skip questions or return to a previous question to change your answer.						
With Regard to the Experienced Environment						
1. How much were you able to control events?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
2. How responsive was the environment to actions that you initiated?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
3. How natural did the interactions with the environment seem?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
4. How completely were all your senses engaged?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
5. How much did the visual aspects of the environment involve you?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
6. How much did the auditory aspects of the environment involve you?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
7. How natural was the mechanism which controlled the camera movement?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
8. How aware were you of events occurring in the real world around you?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
9. How aware were you of the screen and sensors?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
10. How compelling was your sense of objects moving through space?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
11. How consistent was the information coming from your various senses?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
12. How much did your experiences in the virtual environment seem consistent with your real-world experiences?						
NOT AT ALL		SOMEWHAT			COMPLETELY	

13. Were you able to anticipate what would happen next in response to the actions that you performed?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
14. How completely were you able to actively survey the environment using vision?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
15. How well could you fumble in the virtual environment?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
16. How compelling was your sense of moving around in the video environment?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
17. How well could you perceive wind?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
18. To what degree did you feel confused or disoriented at the beginning of breaks or at the end of the experimental session?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
19. How involved were you in the video environment experience?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
20. How distracting wasn't the wind accessory?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
21. How much delay did you experience between your actions and reactions from the simulator?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
22. How quickly did you adjust to the virtual environment experience?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
23. How proficient in moving and interacting with the virtual environment did you feel at the end of the experiment?						
NOT AT ALL		SOMEWHAT			COMPLETELY	
24. How much did the visual display quality not interfere or distract you from performing assigned tasks or required activities?						
NOT AT ALL		SOMEWHAT			COMPLETELY	

25. How much did the control devices not interfere with the performance of assigned tasks or with other activities?

--	--	--	--	--	--	--

NOT AT ALL

SOMEWHAT

COMPLETELY

26. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?

--	--	--	--	--	--	--

NOT AT ALL

SOMEWHAT

COMPLETELY

27. Could you learn the places of the movies in such a way that you would recognise them in the real world?

--	--	--	--	--	--	--

NOT AT ALL

SOMEWHAT

COMPLETELY

28. Were you involved in the experimental task to the extent that you lost track of time?

--	--	--	--	--	--	--

NOT AT ALL

SOMEWHAT

COMPLETELY