Immersive 360º Mobile Video with an Emotional Perspective

João Ramalho
LaSIGE, Faculty of Sciences
University of Lisbon
1749-016 Lisboa, Portugal
+351217500087
joaoramalho90@gmail.com

Teresa Chambel
LaSIGE, Faculty of Sciences
University of Lisbon
1749-016 Lisboa, Portugal
+351217500087
tc@di.fc.ul.pt

ABSTRACT
By appealing to several senses and conveying very rich information, video has the potential for a strong emotional impact on viewers, their sense of presence and engagement. These capacities 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 360º and HD video and metadata and to support more powerful and immersive video user experiences. In this paper, we explore the immersion potential of mobile video augmented with visual, auditory and tactile perceptual sensing with an emotional perspective on the impact on users. Results confirmed advantages in using a multisensory approach to increase immersion and user satisfaction, and identified properties and parameters that worked better in different conditions.

Categories and Subject Descriptors
H.5.1 [Information Interfaces and Presentation (I.7)]: Multimedia Information Systems – *video, hypertext navigation and maps*; H.5.2 [Information Interfaces and Presentation (I.7)]: User Interfaces – *interaction styles, evaluation*;

General Terms
Design, Experimentation, Human Factors.

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

1. INTRODUCTION
Immersion can be defined as the subjective experience of being fully involved in an environment or virtual world, and may be increased by a surround effect, sensory modalities and vividness through resolution [20,5]. Immersion is associated with presence, and the viewer's conscious feeling of being inside the virtual world [20], the perception of self-location [22], and may benefit from realism, e.g. through photo realistic images and spatial audio. Video allows great authenticity and realism, and it is becoming ubiquitous, in personal capturing and display devices, on the Internet and iTV [13,14]. Immersion in video has a strong impact on the viewers' emotions, and especially arousal, their sense of presence and engagement [21]. 360º videos could be highly immersive, by allowing the user the experience of being surrounded by the video. Wide screens and CAVEs with varying angles of projection, possibly towards full immersion, are privileged displays for immersive video view for their shapes and dimensions, but they are not very handy, and especially CAVEs are not widely available. On the other hand, mobile devices are commonly used and represent, by the sensors and actuators they are increasingly incorporating, a wide range of opportunities to capture and display 360ºand HD video enhanced with metadata (e.g. geo-location and speed) having the potential to support more powerful and immersive video user experiences, on their own or as second screens, to be used in complement to wider screens.

In this paper, we explore the immersion potential of mobile interacttive video augmented with visual, auditory and tactile multisensing, through the design and evaluation of new features in Windy Sight Surfers (WSS). This is a mobile application for the capture, visualization and navigation of georeferenced 360º immersive interactive videos (through hypervideo), along trajectories, designed to empower users in their immersive video experiences, both accessing other users' videos and sharing their own. The focus of this paper is on perceptual sensing and its impact on immersion, especially in an increased sense of presence and realism, through the feeling of being inside the video, viewing and experiencing movement speed and orientation, complemented with an emotional perspective, both for video cataloguing and access, and for engagement and satisfaction system evaluation, based on the emotional impact on users. Different conditions were tested, varying: types of video, viewing modes, spatial sound and tactile sensing approaches. Results confirmed advantages in using a multisensory approach to increase immersion, and identified properties and parameters that worked better and are more satisfying and impactful in different conditions that may help to inform design of future mobile immersive video environments.

2. RELATED WORK
The work in this paper builds on our previous work on 360º hypervideo [13], developed for PCs, which evolved to allow capturing, sharing and navigating georeferenced 360º videos and movies, synchronized with maps, and crossing trajectories [14], allowing to 'travel' in other users 'shoes'. Briefly, related work [13,14] concerns to hypervideo and immersive environments (mainly VR and AR images, like Google Street View, seldom video), georenferencing and maps, orientation, cognitive load, and filtering. According to [1], genres more suitable for 360º video from a user perspective include hobbies, sports, or situations with little progress, inviting for exploration. On mobiles, relevant related work concerns to navigation [13], recent PanoramaGL lib for 360º photos, second screens [3], and the use of sensors and actuators, e.g. in art installations where user's movement influences wind (fans) blowing trees [8].

Moon et al. [10] showed that the use of wind output increased the sense of presence in VR, but their application did not allow user interaction, as the movement occurred in a pre-defined path. Cardin et al. [2] presented a head mounted wind display for a VR flight simulator application. Participants determined the wind direction with a variation of 8.5 degrees. Lehmann et al. [7] evaluated the differences between visual only, stationary and head mounted wind, with considerable increase in presence in stationary and head mounted prototypes. But these approaches target VR - not video, nor

ImmersiveMe'13, October 22, 2013, Barcelona, Spain.

http://dx.doi.org/10.1145/2512142.2512144

mobile environments - requiring heavy and very specific equipment. Furthermore, none of them presents methods to capture wind metadata and couple it with video, to increase realism and immersion.

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. Very recently, Merer et al. [9] addressed synthesis and control of sound attributes from a perceptual point of view, based on the characterization of motion evoked by sounds. This concept is not straightforward, involving physical motion and metaphoric descriptions, as used in music and cartoons. They used questionnaires and drawings (associated with control strategies as continuous trajectories, to be used in apps for sound design or music), focusing on shape, direction, size, and speed, and using abstract sounds, for which physical sources cannot be easily recognized. 3D-sound can significantly enhance realism and immersion, by trying to create a natural acoustic image of spatial sound sources within an artificial environment, most development been made by the film industry [4]. Approaches like [12] address sound mapping in virtual environments (VE), with a concern on affective sound design to improve the level of immersion. They use procedural sound design techniques to enhance the communicative and pragmatic role of sound in VE, concluding e.g. that sound walks using headphones produce the most realistic representation of sound because they provide feedback such as distance, elevation, and azimuth. Most of these approaches address audio in general, production of film soundtracks or tend to address virtual reality scenarios, but not scenarios for augmenting immersion in videos.

In the context of eliciting and visualizing user emotions, there are some works, but not so much on video. Mappiness (.org.uk) 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), by prompting users a couple of times a day, on their iPhones. Maps are created marking the places where people have reported feeling happy (orange), very happy (yellow) and extremely happy (green), and provide access to photos posted by the users in these locations. Visch et al. [21] tested the effect of immersion on emotional responses and cognitive genre categorization of film viewers, in two conditions: animated film in 3D (lower) or CAVE (higher immersion). Viewers rated their own emotions and categorized the movies into four basic film genres (action, drama, comedy, and non-fiction), and results showed that stronger immersion led to more intense emotions but did not influence genre categorization. In a previous work [15], we classify and access videos based on the emotions felt by the user while watching them. We explored the use of three biosensors (respiration, heart rate, and galvanic skin response) to detect five of the basic emotions (anger, disgust, fear, joy, and sadness) Ekman identified to be recognized on facial expressions across cultures [6], and represented them with Plutchick's color model [16] in the movie spaces and timelines in the interface.

3. PERCEPTION AND EMOTIONS IN WSS

WSS is an interactive system capable of capturing, publishing, searching, viewing videos and synchronizing with interactive TVs in new ways. During video capture, a smartphone registers several metadata relative to the video, such as the geo-references, speed, orientation, through GPS, and weather conditions (including wind speed and orientation), through the OpenWeatherMap(.org) webservice. After published in the community, videos can be searched by drawing paths on a map, and selecting map areas or filtering criteria. When a video is selected, there are several features designed to increase sense of immersion and presence. In [17], we presented the design and first user evaluation of WSS addressing context awareness and basic sensing experience features with good results and relating context awareness with more control and experience sensing with increased feeling of presence and satisfaction. With the purpose of increasing immersion even further, in this paper we present an approach that builds on this system, integrating new perceptual features, interaction with TVs & Wider screens, and addressing emotion to catalog and search videos and to evaluate users engagement and satisfaction towards an immersive experience.

3.1 Perceptual Features

3.1.1 Visual Perception in 360º Video

Our Research Question 1 (RQ1) was defined as: "Would a full-screen pan-around interface increase the sense of immersion 'inside' the 360º video?". By displaying 360º videos in full screen on an Android tablet, mapping them onto a transitional canvas that is in turn rendered around a cylinder to represent the 360º view, allows the feeling of being "partially" surrounded by the video. Our approaches towards RQ1 consider using the compass within the tablet. By moving the tablet around, users can continuously pan around the 360º video in both left and right directions, as if it was a window to the 360º video surrounding the user. Additionally, users can pan around the video without having to move, as the entire screen consists of a drag interface (Fig. 1).

3.1.2 Tactile Perception through Wind

In this respect, we focused on the Research Question: "Does wind contribute to increasing realism of sensing speed and direction in video viewing?" (RQ2). A Wind Accessory was developed (Figs. 1, 4). This prototype is based on the Arduino Mega ADK; it is mounted on the back of the tablet and controls two fans generating a combined airflow of 180CFM (maximum). While the videos are being reproduced, the fans blow wind to the viewer taking into account the video's metadata. Particularly, the RPM of the fans is relative to the speed of the movement, and the wind's speed and orientation when video was captured, so that, while the camera is facing against the wind orientation, the fans' RPM (rotations per minute) are much higher than when the camera is orientated towards the wind.

Figure 1: Drag interface being used while viewing a 360º video. Wind accessory coupled with the tablet.

3.1.3 Auditory Perception: Spatial Audio

When a video is recorded, the sound is recorded accordingly to the orientation of the camera. To address the question "Does a 3D mapping of the video sound allow for easier identification of the video orientation while it is being reproduced?" (RQ3), the video sound was mapped onto a 3D sound space. JavaScript's Web Audio API (http://www.w3.org/TR/webaudio/) was used to accomplish this, and any standard set of stereo headphones can reproduce the created effect, although high quality headphones are able to increase the realism of the referred effect. In this sound space, the sound source's position is associated to the front angle of the video, 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 are 360º, the sound source's location changes over a virtual circle around the user's head (Fig.2, 4). During user evaluation, users provided feedback on the best values for the virtual "distance" between users' head and the sound sources.

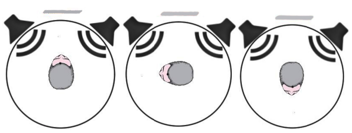


Figure 2: 3D Audio: source changing around the 360º video viewing. Gray stripe on top: trajectory direction.

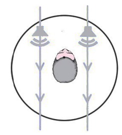


Figure 3: Doppler Effect: Audio changes cyclically as in grey paths.

3.1.4 *Auditory Perception: Cyclic Doppler Effect*

Given the fact that people associate the Doppler Effect to the notion of movement, we experimented to test if "a controlled use of the Doppler Effect can increase the movement sensation while viewing videos?" (RQ4). In order to do so, we added an additional sound layer to the video that cyclically reproduces the Doppler Effect. This sound layer was also implemented using JavaScript's Web Audio API, and therefore the sound corresponding to the Doppler Effect is mapped onto a 3D sound space. The implementation consists of a cycling sound that approaches, passes, and recedes users' head (from the front to the back) (Fig.3). What creates a stronger or weaker movement sensation is the intensity/rate at which the Doppler Effect is reproduced. This rate is set in reproduction time according to the motion speed values stored during video capturing: the higher the speed, the higher the intensity of the Doppler Effect. Regarding the sound used to reproduce the Doppler Effect, several types of sound were experimented and the threshold level of the volume of the sound sources was measured. One of the side effects of this approach might be to try to alert the user for movement when there is no movement, which might become obtrusive rather than beneficial. This problem might be solved with a filter expressing the requirement of a minimum amount of movement in order for the Doppler Effect simulation to execute.

Figure 4: Moving the tablet to view the 360º video. The connected headphones provide a 3D audio space.

3.2 Interaction with TV & Wider Screens

To address the question: "Does the interaction with TVs & wider screens, with video in fullscreen and additional content and navigation control in a 2nd screen, contribute to a more immersive environment?" (RQ5), the application has the capability to interact with wider screens, such as TVs, and take advantage of their screen size. This is a web application running in a standard web browser supporting WebGL, to be shown on a TV, without mouse or keyboard. The interaction is done via the mobile application, as a 2nd screen.

Minimap. When a video is being reproduced in the TV, a minimap with a complete projection of the 360º video is shown to provide increased perception and control of the whole video image (Fig. 5). Overlaid on the minimap, a red rectangle highlights the video angle being reproduced in the TV, which can be changed by dragging the rectangle or by pressing any other part of the minimap. Another way to change the view angle consists in turning the mobile device as if it was a steering wheel.

Geographical Navigation & Orientation in the 360º Videos and Maps. While viewing videos on TV, the mobile application presents a map, which identifies video trajectories (Fig. 5). The path corresponding to the current video is shown in green, while other paths are shown in blue. The trajectory parts of the video that have already been watched are painted red. By touching or dragging the trajectories, the user can navigate to the corresponding video and time.

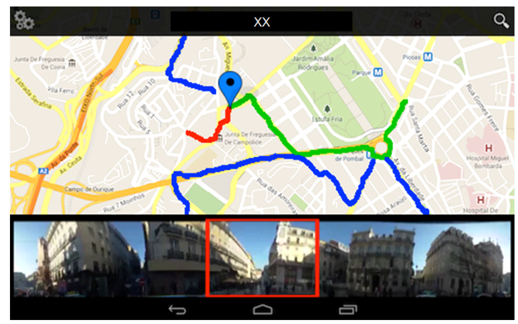


Figure 5: Android application, displaying the map (top) and the minimap (bottom).

3.3 The Emotional Perspective

Striving to provide a more engaging experience to the user, an emotion recognizer system, based on a facial expression recognition framework [11], was used with the application. It is used to recognize the user's emotions when watching video and using the application, through the android device's front-facing camera. It recognizes eight emotions (Neutral, Anger, Contempt, Disgust, Fear, Happiness, Sadness, Surprise), including the six basic emotions identified by Ekman [6]. These results are stored in a database, registering the emotional information associated with users and videos, and are being used in two situations: for emotional video cataloguing and search; and for the evaluation of the environment itself, based on the emotional impact. In the current state of development, our RQ6 in this context is "Do users consider the emotional perspective relevant in the access, search and recommendation of videos?", which will be explored with the following features.

3.3.1 *EmoMap: Emotions on the Map*

The EmoMap feature is associated with the "search through the map" functionality [17], which enables users to search for videos through the selection of areas and paths on a map. With the EmoMap feature, a semi-transparent overlay is added to the map, coloring areas according to the dominant (most frequent) emotions of their videos (areas are defined by having a similar amount of videos). In this approach, we adopted a representation of emotions based on colors, like in the model of Plutchik [16] and our previous work [17] (Fig. 6). This enables the user to identify the map zones that contain more videos associated with specific emotions. Also, users can search for videos associated with specific emotions by expressing the desired emotion in front of the camera of the device, or by filtering through the selection of a list of checkboxes. The search results can be presented in a coverflow list, or in the map.

3.3.2 *EmoMe: My Emotional Profile*

EmoMe is an interface that presents users with personal statistical information regarding their emotional history in the application. In addition to data related to the previously viewed videos, charts will present video trends, as well as recommendations. This feature is being designed and tested to learn about its perceived usefulness, satisfaction and ease of use in order to refine its design and further developments. This recommendation system will take into account data from the user, and from users with similar preferences. As the number of videos watched by the users increases, the preferences and recommendation system is expected to become more precise.

Figure 6: Map with semi-transparent overlay, highlighting dominant emotions of the videos in each area.

3.3.3 Emotional Impact in User Evaluation

To conduct a more representative evaluation of the environment, the emotion recognizer was used as an evaluation tool. In each task, the dominant emotions are logged, which allowed analyzing the user's dominant mood during each stage and during the entire user evaluation process. These results are presented in section 4.2.4.

4. USER EVALUATION

We conducted a user evaluation of WSS's Perceptual features to investigate whether and in which conditions they contribute to a more immersive user experience. We learned about tasks' perceived Usefulness, users' Satisfaction and Ease of use (USE: http://www.stcsig.org/usability/newsletter/0110_measuring_with_use.html). To test dimensions related to Immersion, we used self assessment approaches: the Self-Assessment Manikin, SAM: http://irtel.uni-mannheim.de/pxlab/demos/index_SAM.html, to measure pleasure, arousal and dominance emotions, often associated with immersion; and additional parameters of Presence and Realism (PR) we found relevant. While performing each of the proposed tasks, an emotion recognizer [11] was used in addition, to automatically recognize the most prevalent emotions. We evaluated global immersiveness in the WSS user experience, through a pre and post-event self-assessment Immersive Tendencies and Presence Questionnaires [19,23].

4.1 Method

We performed a task-oriented evaluation based mainly on Observation, Questionnaires and semi-structured Interviews. After explaining the purpose of the evaluation and a short briefing about the concept behind WSS, demographic questions were asked, followed by a task-oriented activity. Errors, hesitations and performance were observed and annotated. At the end of each of the 21 tasks, users provided a 1-5 USE rating, a 1-9 SAM rating and a 1-9 PR rating. Also, the most prevalent emotions identified by the emotion recognizer were logged and comments and suggestions were annotated. Slater states that Presence is a human reaction to Immersion [18]. 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) before the experiment, and an adapted version of the Presence Questionnaire (PQ) after the experiment, with 28 questions each [19,23].

The evaluation had 17 participant users (8 female, 9 male) between 18-34 years old (mean 24). 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 360º videos, and only 6 had heard about 3D audio. The foreseen time for the completion of the 21 tasks was 50 minutes, which was met by all users.

4.2 Results

Results are divided in four subsections, concerning: the perceptual, 2nd screen, and emotional features; the emotional impact of WSS; followed by the Immersive Tendencies and the Presence Questionnaires results and global overall comments.

4.2.1 Perceptual Features

Users were asked to: move around a 360º video by moving the tablet around (T1) and by using the drag interface (T2); and view a 360º video with the wind accessory and identifying the wind direction (T3). Users appreciated the tested features, especially the video navigation by moving the tablet around, which they reported to be a more natural approach when compared to the touch interface; and the wind accessory, which allowed a more realistic sense of speed in video viewing, as the PR results show (T3: PR: 8.9; 8.9), thus confirming RQ2. Despite users favoring the "moving the tablet around" feature for the sense of immersion, the consensus among users was that there are situations where the drag interface can be more suitable, for its flexibility, answering RQ1, and reinforcing the idea that both interfaces are needed and complement each other.

Table 1. USE evaluation of WSS (scale: 1-5).

Features in Task:	**Usefulness**		**Satisfaction**		**Ease of Use**	
	M	σ	M	σ	M	σ
T1 move around	4.8	0.3	4.7	0.3	4.8	0.8
T2 drag interface	4.8	0.4	4.5	0.5	4.8	0.4
Wind Accessory						
T3 wind	4.3	0.6	4.8	0.7	4.9	0.3
3D Audio						
T4 stereo	4.8	0.7	4.5	0.7	4.8	0.5
T5 3D	4.8	1.1	4.7	1	4.8	0.5
Cyclic Doppler Effect						
T6 wind sound	2.7	0.6	2.5	0.5	4.9	0.5
T7 low freq sound	4.4	0.9	4.6	0.8	4.9	0.3
T8 high movement	4.8	0.4	4.7	0.4	4.9	0.4
T9 med movement	4.4	0.4	4.5	0.4	4.9	0.5
T10 low movement	2.8	0.4	3.1	0.4	4.9	0.4
T11 no Doppler	4.7	0.5	4.5	0.6	4.9	0.5
T12 custom Doppler	4.7	0.6	4.7	0.6	4.9	0.3
Interaction with TVs						
T13 navigate TV	4.7	0.3	4.4	0.5	4.8	0.2
T14 drag marker	4.7	05	4.6	0.3	4.8	0.2
EmoMap						
T15 search EmoMap	4.8	2.8	4.9	2.8	4.8	2.8
T16 search checkboxes	4.3	1.4	4.5	1.1	4.9	0.7
T17 search expression	4.6	1.3	4.6	1.4	4.9	0.9
T18 view coverflow	4.3	0.6	4	0.6	4.9	0.3
T19 view map	4.6	0.3	4.6	0.5	4.8	0.3
EmoMe						
T20 charts	4.6	0.3	4.5	0.4	4.9	0.4
T21 recommend view	4.4	0.4	4.4	0.4	4.9	0.3
Overall	**4.4**	**0.9**	**4.4**	**0.7**	**4.9**	**0.5**

In order to test the 3D Audio feature, users were asked to: view a 360º video with headphones and a standard stereo video sound (T4); and view the same video with the 3D sound capability (T5). Regarding T5, users were asked to vary the virtual distance of the speakers (through a slider in the interface) to the users' head and find the optimal distance. In respect to RQ3, 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.

To test the Cyclic Doppler Effect feature, users were asked to view a video with this feature activated, with different sounds in each version. In the first version, a wind sound (T6); and in the second, a low frequency sound (a bass synthesizer sound) (T7). Users needed to choose their preferred sound and were asked to vary the Doppler Effect sound volume and identify the optimal value. Next, users viewed three videos with the Doppler Effect feature activated and were asked to state in which of them they liked this effect the most. The three videos presented situations where the degree of movement was: 1) high (T8); 2) medium (T9); and 3) little/no movement (T10). The order in which the videos were viewed was randomized for each user. Lastly, taking into consideration all the user's preferences regarding the Doppler Effect feature (in T6-T10), users viewed a video twice: once without (T11) and once with (T12) the "custom" Doppler Effect feature, and were asked to state whether they felt this feature increased the movement sensation. Answering RQ4, the SAM and PR values (T12: SAM: 8.8; 8.6; 8.7; PR: 8.9; 9) support that the Doppler Effect increased the movement sensation, although its reproduction must be carefully controlled. Namely, almost all users preferred the low frequency sound, as they stated it was much less obstructive than the wind sound, tending to set the Doppler Effect volume level between 7% and 18% of the main video sound volume. Also, the more movement there is in a video, the more satisfying the Doppler Effect became. Users referred to T8 (high degree of movement) to express a situation where they particularly enjoyed this effect (T8: USE: 4.8; 4.7; 4.9; PR: 8.8; 8.9). On the other side, in videos with no movement, the viewing experience was better without this feature, as the USE and PR results show (T10: USE: 2.8; 3.1; 4.9; PR: 3.9; 3.3). This result confirms the need for the filtering feature that establishes the minimum movement amount for the Doppler Effect feature to be activated.

Table 2. SAM (Pleasure, Arousal, Dominance) and PR (Presence, Realism) evaluation of WSS (scale: 1-9).

Feature in Task:	SAM						PR			
	Pleasure		Arousal		Dominance		Presence		Realism	
	M	σ	M	σ	M	σ	M	σ	M	σ
T1	8.2	0.5	8.7	0.8	8.0	0.7	8.8	0.8	8.8	0.7
T2	7.9	0.6	7.4	0.8	8.8	0.5	8.3	0.9	8.4	0.8
T3	8.3	0.8	8.6	0.9	8.6	0.8	8.9	0.5	8.9	0.5
T4	7.9	0.7	8.1	0.7	8.8	0.7	8.5	0.9	8.4	0.8
T5	8.3	1.2	8.5	1.3	8.6	1	8.7	1	8.8	0.9
T6	3.9	1	6.1	0.8	7.4	1.2	4	1.6	3.7	1.4
T7	8.2	0.9	7.9	0.8	8.4	1	8.2	0.6	8.2	0.5
T8	8.7	1	8.7	0.5	8.5	0	8.8	0.5	8.9	0.6
T9	8	1.7	7.8	0.8	8.1	0.6	8	1.8	7.9	1.5
T10	4.4	0.7	5.2	0.5	6.2	0.4	3.9	0.6	3.3	0.8
T11	7.6	1	7.7	1.3	8.6	1.2	8.3	1.1	8.4	1
T12	8.8	1.2	8.6	0.8	8.7	0.9	8.9	1	9	0.7
T13	8.2	0.7	8.8	0.8	8.3	0.7	8.9	0.5	8.8	0.4
T14	8	1	7.9	0.9	8.1	0.9	8	1.1	8.1	1.2
T15	8.7	0.5	7.8	0.7	8.3	0.5	8.1	0.8	7.8	0.8
T16	7.6	1.9	6.3	1.4	8.8	2.1	7.7	0.9	7.4	0.7
T17	8.7	1.7	8.9	1.5	7.4	2.2	8.3	1	8.5	0.9
T18	8.2	1.1	7.4	0.8	8.2	0.9	8.1	1	7.9	0.7
T19	8.5	1	8	1.1	8.4	1	8.3	1.3	8	0.9
T20	7.8	0.8	7.5	0.9	8.3	1.1	6	0.8	6.1	1
T21	5.9	0.7	5.7	0.9	7.7	0.9	6.1	1.3	6.2	1.1
Overall	**7.7**	**1.0**	**7.7**	**0.9**	**8.2**	**0.9**	**7.8**	**1.0**	**7.7**	**0.9**

4.2.2 Interaction with TV & Wider Screens

While evaluating the interaction with TV & wider screens features, users were asked to connect the mobile application to a TV set, select a 360º video and navigate it through both the "steering wheel" and the minimap displayed on the tablet (T13). Furthermore, users were asked to change to another video by dragging the marker into other trajectory (T14). Answering RQ5, 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 users with a reference to the full 360º angle. Regarding the "steering wheel", users referred that it is very intuitive, but pointed out that its precision must be adjusted individually to each user, 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). Also, very positive comments were directed towards the geographical navigation through the maps, with users indicating it represents an intuitive approach to navigating through the videos.

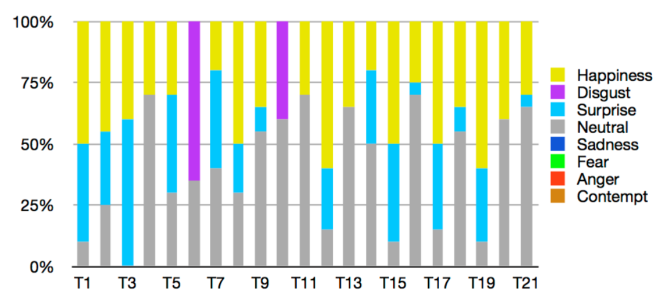
4.2.3 Emotional Features

In order to test the EmoMap feature, users were asked to: find a video with anger as the dominant emotion through EmoMap (T15); and a video with happiness as the dominant emotion through the checkboxes filter tool (T16); and by expressing the desired emotion to the camera (T17); search for neutral videos and view the results in the cover-flow mode (T18); search for neutral videos and view the results in the map (T20). 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 results, users enjoyed the map, stating that the search process was more interesting when they had access to the geo-location dimension. When evaluating the EmoMe feature, users were asked to identify what information was presented in each of the charts (T20); and choose and view a recommended video (T21). All charts were correctly identified and users appreciated the recommendations suggested by the system. Answering our RQ6, users found this emotional perspective interesting in the access, search and recommendation of videos, encouraging further developments in this direction.

4.2.4 Evaluating the Emotional Impact of WSS

Dominant emotions were logged while users performed each assigned task. As figure 7 shows, the user's dominant mood during the several stages of the evaluation was related to positive emotions. In fact, a negative emotion ('Disgust') was only detected in tasks T6 and T10. Task T6 corresponded to the experience of using a wind sound for the Cyclic Doppler Effect, which was not appreciated by the users; T10 corresponded to the experience of using the Cyclic Doppler Effect in a video with no movement, which proved to be a situation where this feature becomes obtrusive rather than an aid.

Figure 7: Recognized Emotions during user Evaluation. (only Happiness, Disgust, Surprise and Neutral were detected)



4.2.5 Global Presence and Immersion Evaluation

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 (table 3 and 4). As Presence is a human reaction to immersion, the PQ scores reveal the global immersiveness of the tested features. Moreover, the improvement from the ITQ to the PQ reveals that WSS surpassed user's immersive expectations of the system.

Table 3: Immersive Tendencies Questionnaire (scale: 1-7).

Tendency to	M	σ
maintain focus on	4.2	1.3
become involved	4.3	1.7
view videos	5.0	1.2

Table 4: Presence Questionnaire (scale: 1-7).

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

5. CONCLUSIONS AND PERSPECTIVES

We presented the motivation and challenges, and described the design and user evaluation, of the sensing, interaction and emotional impact features of WSS towards increased immersive experiences. It is a mobile application that uses a wind accessory, 3D audio, Doppler Effect simulation, and an Emotion Recognition system for the search, visualization and sensing of georeferenced 360º hypervideos, offering the possibility to interact with TVs. The user evaluation showed that the designed features tend to increase the sense of presence and immersion, and that users appreciated them, finding all of them very useful, satisfactory and easy to use. Users showed great interest in the wind device, indicating this is a very effective way to improve the realism of the environment. Using 3D audio was a clear advantage in motion awareness. It is fairly simple to implement and does not require new infrastructure, as any stereo headphones will suffice and the sensors commonly available in mobile devices allow motion detection. According to our tests, the sound sources distance to the user' head in the virtual sound space should be between 1 and 3 meters. The use of the Doppler Effect simulation, when carefully manipulated, as described, can increase the users' movement sensation, especially in videos with a high degree of movement. The emotion recognition appealed to users not only as an innovation, but also as an interesting and useful tool, especially for the video search process, as it introduces a new kind of filtering.

Next steps include: refining and extending our current solutions, exploring further settings for higher levels of immersion, like the CAVE and wide screens, and other modalities to increase users' engagement; exploring 3D audio capture to further increase realism in the spatial audio feature; considering georeferencing, ambient computing and augmented reality scenarios, e.g. as access points to videos shot in the same place at a different time, to compare them 'overlaid', or access videos with similar speed as the current speed experienced by the user (e.g. while traveling on a train).

6. ACKNOWLEDGMENTS

This work is partially supported by FCT through LASIGE Multiannual Funding and the ImTV research project (UTA- Est/ MAI/0010/2009). Special thanks to our project colleagues of [11].

7. REFERENCES

[1] Bleumers, L., Broeck, W., Lievens, B., Pierson, J. (2012) Seeing the Bigger Picture: A User Perspective on 360° TV. In Proc. of EuroiTV'12, 115-124.

[2] Cardin, S., Thalmann, D., and Vexo, F. (2007) Head Mounted Wind. In Proc. of Computer Animation and Social Agents. Hasselt, Belgium, Jun 11-13. 101-108.

[3] Courtois, C., D'heer, E. (2012) Second screen appli-cations and tablet users: constellation, awareness, experience, and interest. EuroiTV'12, 153-156.

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

[5] Douglas, Y., and Hargadon, A. (2000) The Pleasure Principle: Immersion, Engagement, Flow. ACM Hypertext'00, 153-160.

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

[7] Lehmann, A., Geiger, C., Wöldecke, B. and Stöcklein, J. (2009) Poster: Design and Evaluation of 3D Content with Wind Output. In Proc. of 3DUI.

[8] Mendes, M. (2010) RTiVISS | Real-Time Video Inter-active Systems for Sustainability. Artech, 29-38.

[9] Merer, A., Aramaki, M., Ystad, S., and Kronland-Martinet, R. (2013) Perceptual Characterization of Motion Evoked by Sounds for Synthesis Control Pur-poses. In Trans. On Applied Perception, 10(1), 1-23.

[10] Moon, T., Kim, G. J., and Vexo, F. (2004) Design and evalu-ation of a wind display for virtual reality. In Proc. of ACM Symposium on Virtual Reality Software and Technology.

[11] 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.

[12] Namezi, M., and Gromala, D. (2012) Sound Design: A Procedural Communication Model for VE. In Proc. of 7th Audio Mostly Conference, Greece. 16-23.

[13] Neng, L., and Chambel, T. (2010) Get Around 360º Hypervi-deo. In Proc. of MindTrek, Tampere, Finland. 119-122.

[14] Noronha, G., Álvares, C., and Chambel, T. (2012) Sight Surfers: 360º Videos and Maps Navigation. In Proc.of Geo MM'12 ACM Multimedia'12. Nara, Japan. 19-22.

[15] Oliveira, E., Martins, P., and Chambel, T. (2011) iFelt: Accessing Movies Through Our Emotions. In Proc. of EuroITV'2011, Lisbon, Portugal, Jun 29-Jul 1. 105-114.

[16] Plutchik, R. (1980) Emotion: A psychoevolutionary synthesis Harper & Row New York.

[17] Ramalho, J. and Chambel, T. (2013) Windy Sight Surfers: sensing and awareness of 360º immersive videos on the move. In Proc. of EuroITV 2013, Como, Italy. 107-116.

[18] 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, 193-210.

[19] Slater, M. (1999) Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire, Presence 8(5), Oct. 560-565.

[20] Slater, M., & Wilbur, S. (1997) A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. Presence,6, 603-616.

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

[22] Wirth, W., Hartmann, T., Bocking, S., Vorderer, P.,Klimmt, Ch., Schramm, H., et al. (2007) A process model of the formation of spatial presence experiences. Media Psychology, 9, 493-525.

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