Interactive, Multi-device Visualization Supported by a Multimodal Interaction Framework: Proof of Concept

Nuno Almeida[1,2], Samuel Silva[1,2(✉)], Beatriz Sousa Santos[1,2], and António Teixeira[1,2]

[1] DETI – Department of Electronics, Telecommunications and Informatics, University of Aveiro, Aveiro, Portugal

[2] IEETA – Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro, Aveiro, Portugal

{nunoalmeida,sss,bss,ajst}@ua.pt

Abstract. Nowadays, users can interact with a system using a wide variety of modalities, such as touch and speech. Nevertheless, multimodal interaction has yet to be explored for interactive visualization scenarios. Furthermore, users have access to a wide variety of devices (e.g., smartphones, tablets) that could be harnessed to provide a more versatile visualization experience, whether by providing complementary views or by enabling multiple users to jointly explore the visualization using their devices. In our effort to gather multimodal interaction and multi-device support for visualization, this paper describes our first approach to an interactive multi-device system, based on the multimodal interaction architecture proposed by the W3C, enabling interactive visualization using different devices and representations. It allows users to run the application in different types of devices, e.g., tablets or smartphones, and the visualizations can be adapted to multiple screen sizes, by selecting different representations, with different levels of detail, depending on the device characteristics. Groups of users can rely on their personal devices to synchronously visualize and interact with the same data, maintaining the ability to use a custom representation according to their personal needs. A preliminary evaluation was performed, mostly to collect users' first impressions and guide future developments. Although the results show a moderate user satisfaction, somehow expected at this early stage of development, user feedback allowed the identification of important routes for future improvement, particularly regarding a more versatile navigation along the data and the definition of composite visualizations (e.g., by gathering multiple representations on the same screen).

Keywords: Multi-device applications · Multimodal interaction · Interactive visualization

1 Introduction

Human-Computer Interaction has known considerable advances in recent years. The widespread availability of mobile and multimodal devices boosted the proposal of novel interaction modalities and the exploration of multimodal interaction. These new

interaction capabilities, although currently used and explored in different application areas, have not been much considered for Interactive Visualization [1]. Nevertheless, it is of the utmost relevance to explore and understand the strengths and weaknesses of multimodality when used in this context [2], exploring the potential advantages deriving from a richer interaction scenario, allowing adaptability to different contexts [3], and a wider communication bandwidth between the user and the application [4, 5]. In this regard, aspects such as interaction modality choice, adaptability (e.g., different ways of displaying data depending on the hardware or environment), and the combination of modalities assume particular relevance. Furthermore, deriving from the wide range of devices available (smart TVs, tablet, smartphones, etc.), it is also relevant to explore how these may be used to support Visualization [6], whether individually, providing different views, adapted to the device characteristics [3], or simultaneously, providing multiple (complementary) views of the same dataset [7], fostering a richer interaction experience, or as the grounds for collaborative work [8].

One of the application scenarios guiding our efforts in this context is provided by the ongoing Marie Curie IAPP project IRIS[1]. The aim of this project is to provide a natural interaction communication platform accessible and adapted for all users, particularly for people with speech impairments and elderly in indoor scenarios. The particular scenario under consideration, a household, where a family lives (parents, two children and a grandmother), and where different devices exist around the house, and are owned by the different family members, is a perfect match to the challenges identified above. In our view, communication can go beyond the exchange of messages through these media and profit from the dynamic multi-device environment, where similar contents (e.g., a vacation memoir or the family agenda) can be viewed in different manners, adapted to the device and user preferences, and supporting a collaborative interaction effort.

While we have previously presented an approach to a multi-device multimodal application [1], where one user could profit from multiple devices to have complementary views of the same contents, we have yet to explore the use of one application by different users simultaneously, through multiple devices, tackling how each user visualizes contents and interacts, and how each user's interactions are reflected in the overall state of the application.

In line with these ideas, our main goal is to explore multimodal interactive visualization in multi-device settings and the first challenge, addressed in this article, resides on how to best support these features. We do not aim to mimic existing dedicated conference room collaborative systems, where applications are specifically tailored for that purpose. Instead, we want to bring the availability of this kind of features to everyday life devices and applications, enabling its availability in any application.

To that purpose, in Sect. 2 we present related work on multimodal and multi-device applications, in Sect. 3 we consider a W3C based multimodal interaction architecture, in line with our previous work [9–12], and explore its components to serve multimodal interactive visualization. A proof of concept application is then described in Sect. 4 illustrating a set of basic features made possible by the proposed

---

[1] http://iris-interaction.eu.

solution. Section 5 presents the outcomes of a preliminary evaluation, conducted with six participants, to elicit user feedback to guide future efforts. Finally, Sect. 6 presents a brief discussion and conclusions concerning the outcomes and prospective lines of future work.

2   Related Work

A review of recent literature shows several works focusing on multi-display and other multi-device related topics such as ubiquitous multi-device and migratory multimodal interfaces. PolyChrome [13] is a web based application framework that enables collaboration across multiple devices by sharing interaction events and managing the different displays. Another similar solution is the Tandem Browsing Toolkit [14] that allows developers to rapidly create multi-display enabled applications. Conductor [15] and VisPorter [7] are other examples of multi-display frameworks. Thaddeus [16] is a system which enables information visualization for mobile devices.

WATCHCONNECT [17] is a toolkit for prototyping applications that enables interaction through smartwatches. This work presents a different way of interaction that uses the hardware capabilities of smartwatches.

Several works focus on ubiquitous multi-device scenarios, Kernchen et al. [18] explore the processing steps needed to adapt multimedia content and define framework functionalities. HIPerFace [19], from 2011, is a multichannel architecture that enables multimodal interaction and multi-device scenarios, enabling its use in multiples devices.

Other topic related to the use of multimodal and multi-device scenarios is migratory multimodal interfaces. Berti and Paternò [20] describe migratory interfaces as interfaces enabling users to switch between devices while seamlessly continuing their ongoing task. Blumendor et al. [21] describe a multimodal system with several devices, from TVs to smartphones, where the user interface dynamically adapts to the new context and change the used modalities.

Paterno [22] addresses and discusses some aspects that should be considered while designing multimodal and multi-device interfaces.

Shen et al. [23] propose three modes for multi-surface visualization and interaction, namely: independent, reflective, and coordinated. In the first, devices work independently, while in the second each device shows the same content, and in the last it basically shows the same content but from different viewpoints. Alemayehu Seyed [24] presents a study to identify better interaction design for multiple displays, resulting in a set of guidelines to improve user experience.

From this short overview of recent literature we can highlight the community's interest in exploring multimodal interaction in multi-device scenarios, but there seems to be only very few attempts to address it based on existing standards. While the different proposals provide solutions to tackle the required features, their widespread use may be limited by the adoption of specific architectures, in each case. Furthermore, there is no particular focus on how multimodal interaction and multi-device support can be harnessed for interactive visualization.

3   Multi-device Support

This section presents a brief overview of the architectural aspects involved in supporting multimodal multi-device interaction, discussing the main aspects of the adopted multimodal architecture, and briefly describing the devised multi-device approach.

Multimodal Architecture. Our architecture proposal is based on the W3C multimodal architecture recommendations [25] and on previous efforts to create multi-device systems [10].

The W3C standard for multimodal architectures is divided into four modules (see Fig. 1): the interaction manager (IM), responsible to receive all event messages and generate actions; the data model, that stores the information of the IM; input and output modalities, capturing the users' interaction events or presenting information to the user; the runtime framework, the module responsible for the communication between the modules and the necessary services to run multimodal applications.
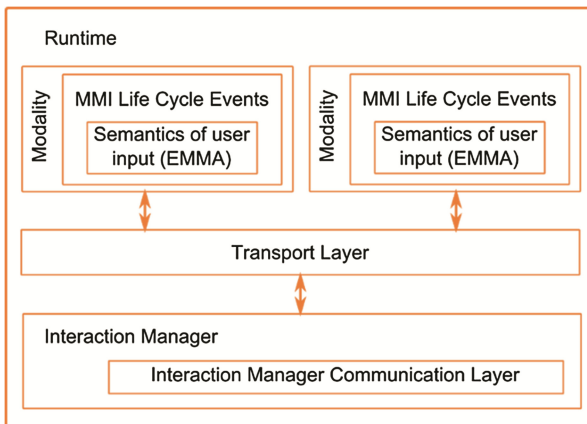

Fig. 1. Multimodal architecture main modules

Going Multi-device. Figure 2 presents the overall architecture of our proposal and possible modalities. Modalities can only communicate with the IM using MultiModal Interaction (MMI) life cycle events [26] carrying EMMA (Extensible MultiModal Annotation markup language) [27], the events information. On the bottom, several classes of devices supported are presented: a computer connected to a large screen, a tablet or a smartphone. Whenever the same modality is connected to the IM, the IM must send a copy of the event to each modality, i.e., interaction is propagated through the different devices and representations.

Aiming for a more ubiquitous approach, we use a cloud based IM capable of managing different modalities in different devices and multiple users.

Each device must run the visualization modality; the touch modality is connected to the visualization modality in order to obtain the objects that the user is interacting with.
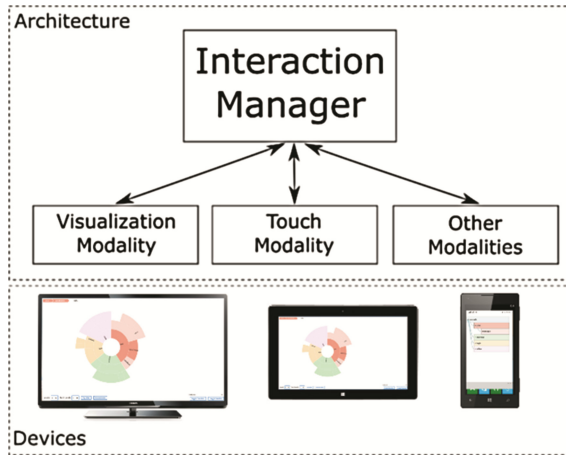
Fig. 2. Architecture and Devices

As a natural outcome of adopting a multimodal architecture, other modalities can be added such as speech [9].

4   Proof of Concept

To support our work and illustrate the capabilities of the proposed approach, we considered a usage scenario extracted from our work on the evaluation of ubiquitous interactive scenarios [28] and created an application prototype to serve as a proof of concept.

Usage Scenario. Dynamic Evaluations as a Service (DynEaaS) is a framework to support the evaluation of multimodal applications in dynamic contexts [28]. Without entering into detail regarding its full range of features, each evaluation session results in data describing all user actions, his/her responses to evaluation tools (e.g., questionnaires) presented during system usage, and all relevant environmental properties and changes. In this context, the considered usage scenario envisages a meeting among three experts to discuss the results of an evaluation session, focusing on the data containing information about the user interaction with a tele-rehabilitation system [28].

The interactions data is organized hierarchically: in the first level are the main components of the application (login, exercise, chat, video and application); in the intermediate levels, subcomponents (e.g., the exercise component has the presentation and list subcomponents); the lower level refers to events and actions (e.g., during exercise presentation there are pause and repeat actions). Each expert has a device capable of running the visualization application (also other modalities can be added to control the application).

Prototype Application. For the development of the proof-of-concept application, the effort was focused in the visualization modality, different modes of visualization were selected based on the data nature. A new modality for the framework was created using

D3.js supporting Interactive Visualization using different representations: the sunburst (Fig. 3a), tree view (Fig. 3b), treemaps (Fig. 3c), and a timeline view (Fig. 3d). Any of these representations can present the same kind of data. The data is organized hierarchically and users can select to focus on a specific level. With a particular focus on the first level, a set of features were added to help users to better understand the data. While moving the mouse over a region a tooltip and a navigation breadcrumb are displayed. This option was considered as opposed to always showing that information as part of the representation since, sometimes, the visualizations encompass large amounts of data and the number of labels would be excessive, becoming difficult to interpret. The number of labels can also be limited by the available screen space and based on the degree of interest of the data they refer to, so that important events are always shown and some labels may be hidden.

a)   b)

c)

d)
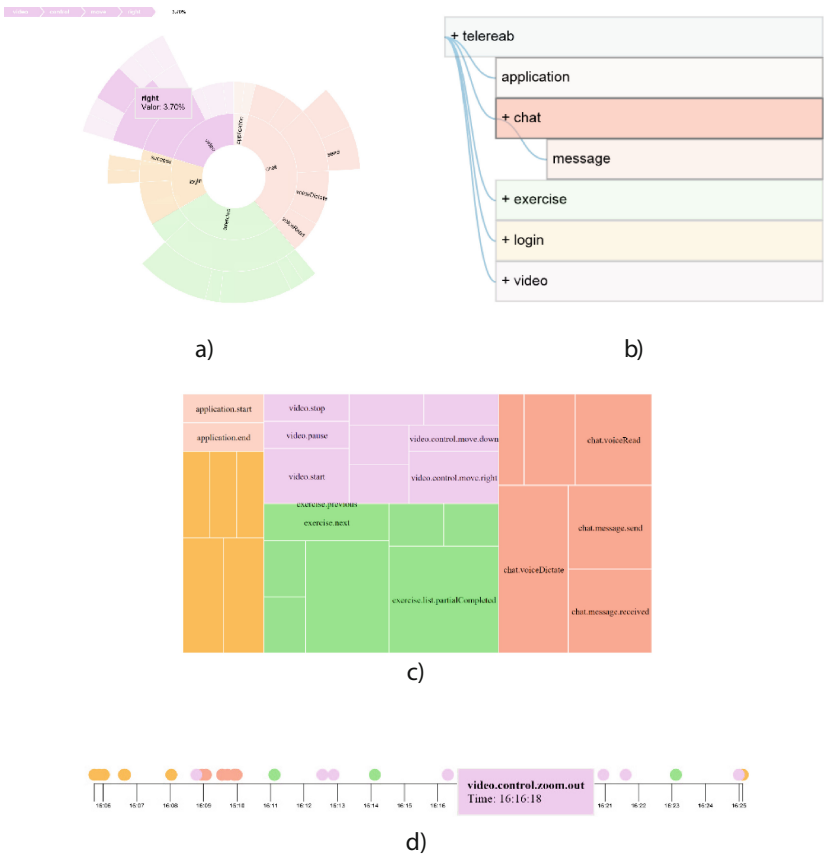
Fig. 3. Data representations available in the prototype application: (a) Sunburst, (b) tree view, (c) treemap, and (d) timeline view.

All devices share a synchronized view of the data, loading the data from the same location. Depending on the device, the modality may default to a representation that best

suits it, depending on various criteria. For example, tree views are used instead of the sunburst for small screen sizes, e.g., smartphones.

5 Preliminary Evaluation

At this point, since we only have a first prototype, serving as a proof-of-concept, our main goal was not to put a strong emphasis on usability results (although not excluding them), since the prototype complexity is still low, and our main concern was to provide a basic set of technical features. Therefore, we were particularly interested in performing a preliminary formative evaluation that could elicit user feedback and suggestions, yielding requirements to guide further developments. The study was conducted with 6 participants, all male, aged between 25 and 35 years old.

5.1 Method

Based on Pinelle et al. [29], we created a plan to evaluate the prototype's usability. First, the system was explained to the users. Then, users were asked to complete two sets of

Table 1. Evaluation tasks

Individual tasks	
Task 1	Find which of the components was most used
Task 2	Find if the user made a mistake dictating or the recognition didn't worked well
Task 3	What was the total time of the session? Compare with the time the user took to perform the exercises
Task 2 (Tablet)	Select the exercises events and verify if the user concluded every exercise.
Task 3 (PC)	Change the view to see each individual event, filter the chat events and observe the time when the user received messages. When were the messages received?
Task 4 (Smartphone)	Select each event until "exercise.Presentation". What is the percentage?
Task 5 (Tablet)	View all names in the visualization.
Task 6 (Smartphone)	Select video.control. What was the most used control?
Group tasks	
Task 1 (PC)	Compare the number of interactions between the video control and chat control. What is the value of each?
Task 2 (Tablet)	Select the exercises events and verify if the user concluded every exercise.
Task 3 (PC)	Change the view to see each individual event, filter the chat events and observe the time when the user received messages. When were the messages received?
Task 4 (Smartphone)	Select each event until "exercise.Presentation". What is the percentage?
Task 5 (Tablet)	View all names in the visualization.
Task 6 (Smartphone)	Select video.control. What was the most used control?

tasks, as described in Table 1. The first set of tasks, was to be conducted once, individually, using a single device, while the second set should be performed in group, with each user working on a different device (PC, tablet or smartphone). In the second task set, each user had his/her own task, but others could also interact to find the result faster.

A subjective evaluation approach was considered, in which users were observed performing the tasks, incidents were registered, and users were encouraged to think aloud. In the end, users were asked to fill a questionnaire, based on the System Usability Scale (SUS) [30]. The scale goes from one to five where one is strong disagreement and five strong agreement. Furthermore, using the same scale as the SUS, other items were added to the questionnaire (Table 2) to analyse the users' preferences concerning the visualizations and their usage in multi-device contexts. Also, users were asked to order visualizations according to their preferences.

**Table 2.** Questions added to SUS, answered in the same scale

Q1	Different visualizations helped to better understand the data?
Q2	Different visualization helped to navigate through the data?
Q3	It is easy to get information from the "sunburst"?
Q4	It is easy to get information from the "timeline"?
Q5	It is easy to get information from the "treemap"?
Q6	The "breadcrumb" helps to locate the information?
Q7	The "Tooltips" and highlight helps to locate the information?
Q8	Combining visualizations helped to understand the information?
Q9	The smartphone is helpful in this context?

5.2 Results

The calculated SUS score was 58 %. While the score was not a great result, this was somehow expected since our main focus, at this stage, was on a first prototype including all the basic technical features supporting the multimodal multi-device interactive visualization. Nonetheless, the other evaluation methods allowed to identify the users' difficulties and retrieve suggestions. Users had some difficulties understanding the data at first since they were not acquainted with the specificities of the application from where the data were retrieved. They always looked for the information using the predefined visualization when, for instance, in the second task, they needed to change to the timeline, which is a complementary visualization, to obtain the results. Users also showed some difficulties finding how to select a different visualization. Also, they struggled to find an event in the timeline. Most of these difficulties were in the first set of tasks, where tasks were individual, and the participants were using them for the first time. In the second set, they were able to communicate and help each other finishing the tasks.

Figure 4 presents the results of the questionnaire. In the users' opinion, the treemap visualization and breadcrumb did not help much. On the other side, the sunburst and the timeline, as well as the tooltips, helped understand the data. Users found the possibility of having different visualizations and the use of the smartphone helpful.
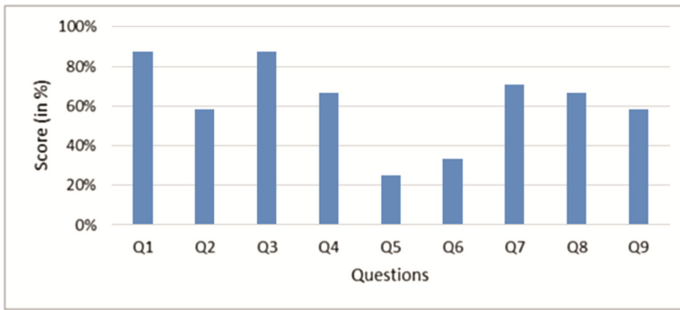
Fig. 4. Overall results obtained from the questionnaire. Please refer to Table 2 for the considered questions.

Resulting from the 'think aloud' use of the prototype, several interesting suggestions were gathered, such as:

Provide a way to differentiate error events from general events
Be able to display the sunburst and timeline on the same screen
Zoom in the timeline in the horizontal
Use the breadcrumb to navigate to previous levels

6   Conclusions

In this first stage of our work, we show how a multimodal architecture, adopted to support multimodal interaction, can also easily encompass the features needed to support multi-user, multi-device interactive visualization. A proof of concept application shows how the visualization modality can work, enabling users to simultaneously interact with the same data and entities while choosing their own representation preferences in the context of the used device. A preliminary evaluation of the application prototype has been carried out to assess the users' overall opinion regarding the provided features (e.g., different representations and synchronous functioning among devices, possibly using different representations for each device) with positive outcomes and ideas for further work.

By taking advantage of a multimodal framework to provide the multi-device features, we are also potentially bringing visualization into multimodality. At its current stage, apart from the visualization modality, the presented proof of concept still does not explore multiple modalities in service of visualization. Nevertheless, inherent to the features of the adopted architecture, a speech synthesis based output modality, for example, would be easy to add [9] along with gaze, as we recently showed for another application domain [31, 32]. This obviously does not mean that innovative approaches to interactive visualization appear automatically, but that the technical effort to add support for those modalities is considerably reduced, leaving room for their creative use in service of visualization, a path we will continue pursuing.

Addressing how the visualization adapts to the characteristics of the data and device is also one of our current lines of work, in line with the proposal of generic interaction modalities aligned with the MMI architecture standard (e.g., for speech interaction [9, 33]).

References

1. Lee, B., Isenberg, P., Riche, N.H., Carpendale, S.: Beyond mouse and keyboard: expanding design considerations for information visualization interactions. IEEE Trans. Vis. Comput. Graph. **18**, 2689–2698 (2012)
2. Ward, M.O., Grinstein, G., Keim, D.: Interactive Data Visualization: Foundations, Techniques, and Applications. CRC Press, Natick (2010)
3. Roberts, J.C., Ritsos, P.D., Badam, S.K., Brodbeck, D., Kennedy, J., Elmqvist, N.: Visualization beyond the desktop–the next big thing. IEEE Comput. Graph. Appl. **34**, 26–34 (2014)
4. Jaimes, A., Sebe, N.: Multimodal human-computer interaction: a survey. Comput. Vis. Image Underst. **108**, 116–134 (2007)
5. Lee, J.-H., Poliakoff, E., Spence, C.: The effect of multimodal feedback presented via a touch screen on the performance of older adults. In: Altinsoy, M., Jekosch, U., Brewster, S. (eds.) HAID 2009. LNCS, vol. 5763, pp. 128–135. Springer, Heidelberg (2009)
6. Schmidt, B.: Facilitating data exploration in casual mobile settings with multi-device interaction (2014). http://elib.uni-stuttgart.de/opus/volltexte/2014/9420
7. Chung, H., North, C., Self, J.Z., Chu, S., Quek, F.: VisPorter: facilitating information sharing for collaborative sensemaking on multiple displays. Pers. Ubiquitous Comput. **18**, 1169–1186 (2014)
8. Isenberg, P., Elmqvist, N., Scholtz, J., Cernea, D., Ma, K.-L., Hagen, H.: Collaborative visualization: definition, challenges, and research agenda. Inf. Vis. **10**, 310–326 (2011)
9. Almeida, N., Silva, S., Teixeira, A.: Design and development of speech interaction: a methodology. In: Kurosu, M. (ed.) HCI 2014, Part II. LNCS, vol. 8511, pp. 370–381. Springer, Heidelberg (2014)
10. Almeida, N., Silva, S., Teixeira, A.J.S.: Multimodal multi-device application supported by an SCXML state chart machine. In: Workshop on Engineering Interactive Systems with SCXML, the Sixth ACM SIGCHI Symposium on Computing Systems (2014)
11. Almeida, N., Teixeira, A.: Enhanced interaction for the elderly supported by the W3C multimodal architecture. In: Proceedings of 5ª Conferência Nacional sobre Interacção (2013)
12. Teixeira, A.J.S., Almeida, N., Pereira, C., e Silva, M.O.: W3C MMI architecture as a basis for enhanced interaction for ambient assisted living. In: Get Smart: Smart Homes, Cars, Devices and the Web, W3C Workshop on Rich Multimodal Application Development. Metropolitan Area, New York (2013)
13. Badam, S., Elmqvist, N.: PolyChrome: a cross-device framework for collaborative web visualization. In: Proceedings of Ninth ACM International Conference on Interactive Tabletops and Surfaces (2014)
14. Heikkinen, T., Goncalves, J., Kostakos, V., Elhart, I., Ojala, T.: Tandem browsing toolkit: distributed multi - display interfaces with web technologies, pp. 142–147 (2014)
15. Hamilton, P., Wigdor, D.J.: Conductor: enabling and understanding cross-device interaction. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI 2014, pp. 2773–2782. ACM Press, New York (2014)

16. Woźniak, P., Lischke, L., Schmidt, B., Zhao, S., Fjeld, M.: Thaddeus: a dual device interaction space for exploring information visualisation. In: Proceedings of the 8th Nordic Conference on Human-Computer Interaction, pp. 41–50 (2014)

17. Houben, S., Marquardt, N.: WATCHCONNECT: A toolkit for prototyping smartwatch-centric cross-device applications. In: Proceedings of 33rd Annual ACM Conference on Human Factors in Computing Systems (2015)

18. Kernchen, R., Meissner, S., Moessner, K., Cesar, P., Vaishnavi, I., Boussard, M., Hesselman, C.: Intelligent multimedia presentation in ubiquitous multidevice scenarios. IEEE Multimed. **17**, 52–63 (2010)

19. Weibel, N., Oda, R.: Hiperface: a multichannel architecture to explore multimodal interactions with ultra-scale wall displays. In: ICSE 2011: Proceedings of the 33rd International Conference on Software Engineering (2011)

20. Berti, S., Paternò, F.: Migratory multimodal interfaces in multidevice environments. In: Proceedings of 7th International Conference Multimodal interfaces. ACM (2005)

21. Blumendorf, M., Roscher, D., Albayrak, S.: Dynamic user interface distribution for flexible multimodal interaction. In: International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction on - ICMI-MLMI 2010. p. 1. ACM Press, New York (2010)

22. Paterno, F.: Multimodality and multi-device interfaces. In: W3C Workshop on Multimodal Interaction, Sophia Antipolis (2004)

23. Shen, C., Esenther, A., Forlines, C., Ryall, K.: Three modes of multisurface interaction and visualization. In: Information Visualization and Interaction Techniques for Collaboration Across Multiple Displays Workshop associated with CHI (2006)

24. Seyed, A.: Examining user experience in multi-display environments (2013)

25. Dahl, D.A.: The W3C multimodal architecture and interfaces standard. J. Multimod. User Interfaces **7**(3), 171–182 (2013)

26. Bodell, M., Dahl, D., Kliche, I., Larson, J., Porter, B., Raggett, D., Raman, T., Rodriguez, B.H., Selvaraj, M., Tumuluri, R., Wahbe, A., Wiechno, P., Yudkowsky, M.: Multimodal architecture and interfaces: W3C recommendation (2012)

27. Baggia, P., Burnett, D.C., Carter, J., Dahl, D.A., McCobb, G., Raggett, D.: EMMA: Extensible multimodal annotation markup language (2009)

28. Pereira, C., Almeida, N., Martins, A.I., Silva, S., Rosa, A.F., Oliveira e Silva, M., Teixeira, A.: Evaluation of complex distributed multimodal applications: evaluating a telerehabilitation system when it really matters. In: Zhou, J., Salvendy, G. (eds.) ITAP 2015. LNCS, vol. 9194, pp. 146–157. Springer, Heidelberg (2015)

29. Pinelle, D., Gutwin, C., Greenberg, S.: Task analysis for groupware usability evaluation. ACM Trans. Comput. Interact. **10**, 281–311 (2003)

30. Lewis, J.R., Sauro, J.: The factor structure of the system usability scale. In: Kurosu, M. (ed.) HCD 2009. LNCS, vol. 5619, pp. 94–103. Springer, Heidelberg (2009)

31. Vieira, D.: Enhanced multimodal interaction framework and applications. Master thesis, Aveiro, Universidade de Aveiro, (2015)

32. Vieira, D., Freitas, J.D., Acartürk, C., Teixeira, A., Sousa, L., Silva, S., Candeias, S., Dias, M.S.: Read That Article: Exploring synergies between gaze and speech interaction, pp. 341–342 (2015)

33. Almeida, N., Teixeira, A., Rosa, A.F., Braga, D., Freitas, J., Dias, M.S., Silva, S., Avelar, J., Chesi, C., Saldanha, N.: Giving voices to multimodal applications. In: Kurosu, M. (ed.) Human-Computer Interaction. LNCS, vol. 9170, pp. 273–283. Springer, Heidelberg (2015)