Aspects of Voice Interaction on a Mobile Augmented Reality Application

Tiago Araújo[1](✉), Carlos Santos[1], Brunelli Miranda[1], Nikolas Carneiro[1], Anderson Marques[2], Marcelle Mota[1], Nelson Neto[1], and Bianchi Meiguins[1]

[1] Universidade Federal do Pará (UFPA), Belém, Brazil
tiagodavi70@gmail.com, gustavo.cbcc@gmail.com,
brunelli.miranda@gmail.com, nikolas.carneiro@gmail.com,
cellemota@gmail.com, dnelsonneto@gmail.com,
bianchi.serique@gmail.com

[2] Universidade Federal Rural da Amazônia, Capanema, Brazil
andmarques2006@gmail.com

Abstract. Mobile Augmented Reality has become more popular mainly because computational resources available in mobile devices, and in the enhanced view of real world that can be seen by the user. The interaction becomes an important point for success of these applications, featuring a natural and intuitive way for the user, and the chance of one, or two hands free for other interaction or activity. Therefore, this work presents the usability analysis of a Mobile Augmented Reality, the ARGuide, with the use of a voice service for interaction. The usability test analysis is based upon the most common interaction tasks of the users in this type of application. In the end, some good practices for the application interface building and speech interaction are shown.

Keywords: Mobile augmented reality · Voice interaction · Speech recognition · Arguide

1 Introduction

Technological development and the rising popularity of Mobile Augmented Reality (MAR) are reasons to widespread usage of this technology. Usability studies of MAR applications can help developers improve user experience, consolidating the use of these applications on several areas. Martínez et al. [1] shows some challenges for MAR application development, and among the presented challenges we can highlight the shortage of development patterns and little space for showing information. An alternative for present and organize information in MAR applications is using natural language.

Natural language interaction is the communication between human and machine using a language familiar for human [2]. This sort of interaction is important due to benefits provided to the user, among them, actions with higher intuitive interaction, minimizing cognitive effort and allowing the user concentrate in the task, instead of in the interaction [3]. Interaction by voice commands is an alternative for use with MAR applications, since tablets and smartphones have a built-in microphone.

S. Lackey and R. Shumaker (Eds.): VAMR 2016, LNCS 9740, pp. 199–210, 2016.
DOI: 10.1007/978-3-319-39907-2_19

It is common for mobile applications to offer many features to the user, reflecting on complex menus. These menus can hide some features, consuming user's time in the search of them. In this scenario, voice interaction can ease the search of some feature or set a shortcut for it [4].

The usage of voice in tablets and smartphones provides another advantage for the user. To hold the device with one hand and interact with the free one can be uncomfortable, as the device is bigger and heavier, like tablets. Interactions by voice commands can inhibit this problem, making the user interact with the application even holding the device with both hands.

In this work, a MAR application received a voice recognition service to help in navigation and content visualization using Points of Interest (POI). We analyze aspects of voice interaction and speech recognition based on a usability evaluation. We focus on task unit of the tests, like the navigation on a map and selection of Points of Interest in a MAR application. The application used for this work is the ARGuide [5].

2 Background

2.1 Voice Interaction

Speech is one of the most used form of communication of human being, standing between the three forms of communication (voice, gestures and facial expression) most used in everyday life [6]. The voice, when used in an application, is inserted in natural language concept.

Natural language interaction is the communication between human and machine using a familiar language for the human [2]. The Human Computer Interaction (HCI) field assigns an important role for this sort of interaction, due its benefits for the user, naming, actions that are more intuitive, minimizing cognitive effort and allowing the user to focus in the task, and not in the interaction [3].

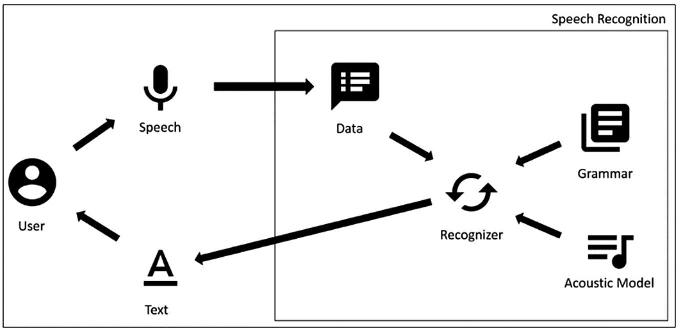

Fig. 1. Default model of a generic speech recognition service

Speech usage in smartphones and tablets also presents these advantages for the user. It is increasingly common that mobile applications developers offer more functionalities to the user, but this rise on functionalities requires more complex menus, which can hide some functions from the user, or demand a long time to the desired function to be found. In this scenario, voice interaction can ease the search of the user for a function, or even use it as a shortcut for the function [4]. Generally, speech recognizer has a default model, even varying little from application to application. Figure 1 has a diagram to show this model.

The recognizer starts when the user commands with the speech, this speech converts in data for interpretation, and is processed with the applications acoustic model, to interpret the signal, and with the grammar, to confirm if the speech is within the applications choose of words. Voice interaction presents practical use in various scenarios, but there few studies that use it with MAR.

2.2 Mobile Augmented Reality

MAR is the use of Augmented Reality (AR) concepts in mobile platforms, which blend merge the virtual content and real environment scenes in the mobile device screen [7]. This is done applying a layer of virtual information above a real scene.

Commonly the MAR applications show two main approaches to blend virtual content in the real scenes: markers or location. The marker approach uses recognition software of a particular pattern (such as QR Code) to select content, and the location approach uses the device's GPS and camera orientation to define what the user sees. The main characteristics of MAR are [8]:

- Presentation of augmented environment in mobile devices screen;
- Track technology: GPS, and pattern recognition of images;
- The graphical system is responsible for virtual objects rendering in augmented scenes;
- The worlds blend system mix real world and virtual objects.

The challenges in MAR application development are [9, 10]:

- Device sensors integration,
- Low precision in track technology;
- Limitations and divergent devices characteristics;
- User's interface variance;
- Lack of development patterns for MAR applications;
- Energy cost

Some guidelines must be followed to MAR applications development [11, 12]:

- Consider target user's profile, outdoor use, one or both hands and time of applications usage;
- Follow the good practices of AR usability and mobile applications. Leave the augmented scene clean, big icons and fonts, layers, 3D object interaction and consider real world;
- Consider devices limits, brightness, light reflection and low precision tracking;
- User's perception and cognition must be stimulated;

- Presentation of virtual information, consider information amount, information representation, information place and multiple visions;
- Evaluation: Adapt guidelines of usability evaluation for AR and mobile applications and evaluate a field at once using real users.

2.3 Related Applications

In order to support and construct the application's functionality in this paper, we selected three applications that based the choices of interaction and GUI design of the proposed application.

ARCity [13] is a MAR application that helps users to find a city's tourist attractions. This app has an integrated map and a classic RA browser of MAR applications. The ideas of do not leave the application to show the route in the map was well evaluated and is available in this article's application.

Immersive Tour Post [14] is a system that uses immersion with audio and video in an important or historic place of a city. The device is located in front of a POI and the tourists can know the history about the tourist attractions with an immersive video. The idea of the POI immersion is the great advantage of this technology, so the application proposed in this paper uses this concept in RA browser, although the user is not fully immersed in the application, this purpose is reached when user interact with the rotating POIs through its own axis.

Time Machine [15] is a MAR application that implements the panoramic image concept to make a list of tourist attractions in the current state and its past. The user can follow a visual timeline of these sights. The proposed application shows this timeline with historical photos of POIs. The images describe the story of these points with respect to what is depicted in the image.

3 ARGuide

ARGuide is a MAR application that uses georeferenced content as base for user navigation. The application allows the user to explore POI changing coordinated visions to access available content. The associated content of each POI can be multimedia, and the visions are different forms to visualize a POI. Coordinated visions means that the changes made in one vision reflects in all other, allowing that the user notices the same information with different perspectives. The available visions on application are:

- Map: Visualization of POIs application in a geographical map.
- RA Browser: Using the POI position, sets a marker in application camera correspondent to real world position
- POI List: Organizes POIs in a list.
- POI View: Shows POI content in a window, arranging text and media available.

The speech recognition service mainly affected two aspects of ARGuide, the GUI and architecture.

3.1 Graphical User Interface

The original ARGuide's GUI was changed to improve voice interaction, since previous ARGuide's interfaces were not developed for supporting voice interaction. The main changes of GUI are in Fig. 2.

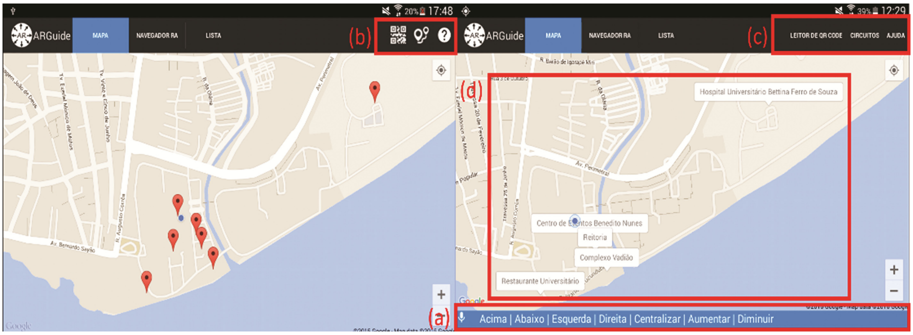
17:48
ARGuide
MAPA
NAVEGADOR RA
LISTA
(b)
12:29
ARGuide
MAPA
NAVEGADOR RA
LISTA
(c)
LEITOR DE QR CODE
CIRCUITOS
AJUDA
(d)
Hospital Universitário Bettina Ferro de Souza
Av. Perimetral
Centro de Eventos Benedito Nunes
Reitoria
Complexo Vadião
Restaurante Universitário
Av. Bernardo Sayão
(a)
Acima | Abaixo | Esquerda | Direita | Centralizar | Aumentar | Diminuir

Fig. 2. ARGuide's GUI in Map view, highlighting the main changes of the application

The changes seen in Fig. 2 includes:

- Use of text labels instead of visual representations. Labels in 2(a) inform the user the commandos available to the view in use, changing when the user changes the context (the view) in order to reflect controls of that view. Labels in 2(c) replaced icons in 2(b), in order to easy the interaction by suggesting to the user the commands that can be spoken.
- The POIs representation in the map were also changed to better suite voice interaction. In place of red markers, POIs are represented by text boxes with its name, allowing user to speak the POI name to open. POIView was adapted including a number for each media available for that POI, so the user can say the number of the media that should be opened.

3.2 Voice User Interface

To assess speech interaction, users should be able to navigate the ARGuide only using voice commands and with all the functionalities available by touch. The speech recognizer is always listening for user commands, starting and stopping with the application.

To allow speech interaction the ARGuide had gone through an architectural refactoring. A voice service has a link to a module of application as shown in Fig. 3. The application must set the actions to match the words recognized and execute them accordingly.

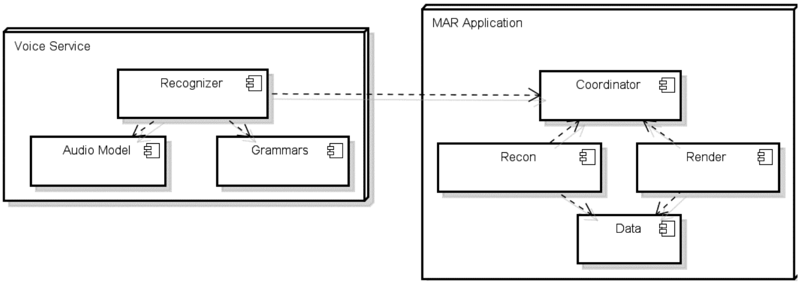

Fig. 3. Architecture for MAR application and voice recognition service

The Voice Service package are made of three components in order to process voice signals and extract action to the application. The Recognizer access the microphone and transforms its data to the useful information for the Audio Model, that process the signal and returns the recognized word (or sentence). The returned word is compared to the application grammar and if it is valid, the Recognizer sends it to the MAR Application package for processing the interaction. This package compares the received word with the active context in the application, and then execute the action (that depends both on the word and on the active context).

The grammar words are defined according to [4], the interaction follow guides from [11, 12] for MAR applications. Selected words should also be recognized by near variations, as plurals, synonyms and gender variations. Table 1 defines some interaction on the tool.

Table 1. Relations between voice commands, actions done by it and its GUI correspondent.

Voice commands	Action	GUI correspondent
esquerda, direita, cima, baixo (left, right, up, down)	Pan	Map /Image
aumentar, diminuir (zoom in, zoom out)	Zoom	Map /Image
aumentar, diminuir (zoom in, zoom out)	Ampliar ou diminuir raio de alcance.	AR Browser
subir, descer (move up, move down)	Scroll	List
qrcode, circuitos, ajuda, mapa, navegador ra, lista (qr code, circuits, help, map, ra browser, list)	Change between Menus & Views	QR Code reader, Circuits, Help and Views
<POI Name > [a]	Select POI	Map Marker | AR Marker | List Item

[a] This comand is dynamic and is dependable of the database items of the application.

The choice of the grammar words aimed to make possible that all the navigation are made using only the voice. For map navigation and image manipulation the same set of words were used due to they are analog to the actions they trigger. For the AR Browser the word set are the same as the zoom controller in map and image manipulation, due to the analog action are the same. The POI List only has scrolling controllers for going up and down.

Some commands recognized in every context, those are: menu selection, once the menus are always visible in the application; circuit exchange, that filter visible POIs in all views; help, that changes from view to view; view swap; POI selection, via POI name (except from the QR Code Reader).

4 Evaluation

Two evaluation methods were used to assess the interface and the voice commands in the application. One of those methods are based on tasks, which recorded interactions of each participant, and the other method is the usage of questionnaires, where the user answered questions about the application use and the user profile.

Table 2. List of tasks for the user test, divided by task units and classified by its complexity

Task number	Task	Task unit	Complexity
1.	Select a Circuit	a. Select circuit menu	2-Medium
		b. Select circuit	
2.	Find a POI using AR Browser	a. Select AR Browser view	3-Hard
		b. Fit AR view i) Apply zoom	
		c. Select POI	
3.	Find a POI using map	a. Select Map View	2-Medium
		b. (Optional) Fit Map view i) Use pan/zoom	
		c. Select POI	
4.	Find a POI using list	a. Select List View	2-Medium
		b. (Optional) Roll list	
		c. Select POI	
5.	Use QR Code Reader to find a POI	a. Select QR Code View	1-Easy
		b. Scan QRCode*	
6.	Navigate in a picture using pan and zoom	a. Select a picture	2-Medium
		b. Use pan/zoom	

4.1 Test Definition

Laboratory tests were conducted with 10 participants with low noise level (low ar-conditioned noise). These users received training before using the app, through a show-case that presented app's functionalities available by touch and one POI selection via voice. The tasks used in this test are described in Table 2. The "Task" column defines the tasks proposed to the user and "Task Unit" defines the steps that compose each task.

We take complexity of a task by the minimum number of interactions (commands) needed in order to complete the task, one command is an easy task, two commands are medium and three commands make a hard task.

These tasks cover the main functionalities in the application for testing if the interface is proper for voice command interaction and if the chosen commands are meaningful and enough for the interaction. The following question were made to the user:

- How hard was the task?
- Have the application interface helped you completing the task?
- Was the speech recognition satisfactory for this task?

These questions allowed assess if the application interface helps the user completing tasks and if voice interaction helped in the same goal. The profile questionnaires contained the following questions:

- What is your age?
- How frequently do you use smartphones?
- Have you used speech recognition interaction in your smartphone before?
- Have you ever used a Mobile Augmented Reality application before?
- What have you liked in the application?
- What have you not liked in the application?
- Would you use the application if it were available for download in an app store?
- What would you point as positive and negative points in the application?
- In general, what is your evaluation of the application?

5 Results

The evaluation results made are based on questionnaire answers and tests video analysis. The data extracted from the videos is the number of errors related to speech recognition in each task unit and the number of errors related in which the GUI was not clear to the user and not lead him to end of the task. For the user profile, Fig. 4 binds two graphs with the users' answers of questionnaires.

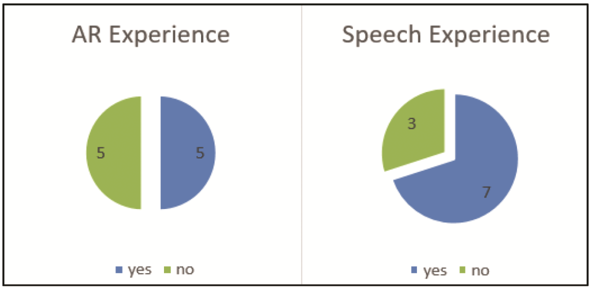

Fig. 4. Answers for user profiles, defining AR and speech experience

Most users has experience with voice interfaces e half of them used an AR application at least once. Age range of the users is between 22 and 42 years old and all of them use smartphone daily. The users' familiarity with the technology used can be an important factor to fulfillment of all tasks. There is not failure in any of them. Figure 5 shows the error average for the tasks of the users, divided by its respective task units.

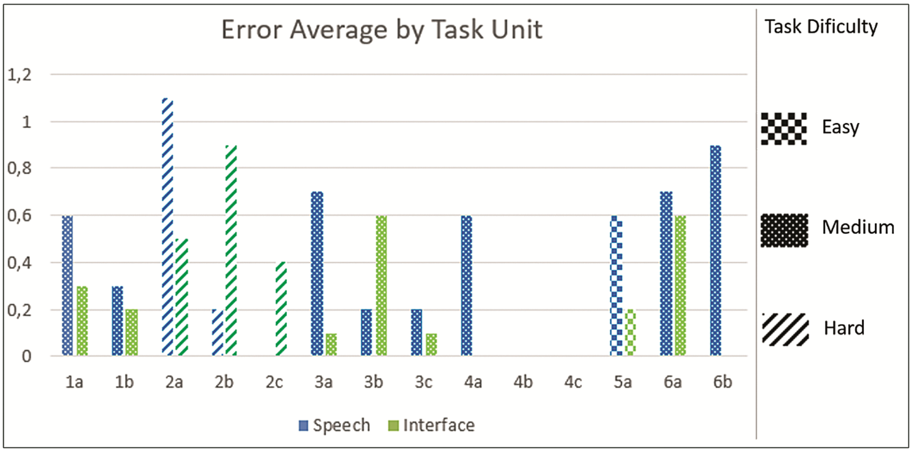

Fig. 5. Error average of user test, divided by task unit

The task with the higher error average for speech recognition is the task unit 2a, to select the AR Browser view from Map view. The word composition from this command ("Navegador RA"-"AR Browser") is divided by two, being the second word two spelled letters. The recognition system was not able to recognize this composition properly for all users, and most users opted to use the single word "navegador" ("browser"), supported as synonymous for the same action, as the voice command. Same difficulty can be seen in task 5a, swap to the QR Code Scanner view, that asks for a single voice command, but has a high error rate of speech recognition compared to error average of GUI.

The task with the higher error average for GUI is the task unit 2b, to extend the range radius of AR Browser. The users should apply zoom and extend the radius to see the marker of the POI. Most users had trouble to identify the widget to extend the range, and some had to remove all range from the AR Browser to understand that they should extend the range to see further POI.

Bugs found in the application by the users through the test, and one of them had influence in the map navigation. One of the words for pan the map was not being recognized ("acima"-"upward"), and the users had to use a synonymous ("cima"-"up") to do the navigation. A context bug, when changing POIView and QR Code Scanner, generated some confusion for the users, preventing them to navigate directly from the POIView to QR Code Scanner. Two users indicated another bug, they noticed that choosing a media in POIView after coming from the List view, the selection would choose the item list instead of the media. Problems reported by the users in questionnaire and the problems found analyzing the videos are listed below:

- Microphone icon in the subtitle looks like a button to start voice interaction (Fig. 2(a)).
- Some words on plural can not be recognized like the ones in the singular, and the contrary sometimes is true too.
- The tabs do not indicate that the view can be changed by voice.
- The AR Browser uses the camera in the same way that the QR Code Scanner, and some users confused its use.
- Circuit change is not clear for the user.
- For the map, the voice commands simulated pan with touch gestures, when yout move for a side and the map pans for the other side. This interaction troubled some users, and was well received by others.

A POI outside a circuit could not be selected, even if the user knows its name. Two users expressed at the end of the questionnaires that the although the application has various views, it keeps the same commands for some actions, like swap views and select POI, causing them to learn, not decorate the commands. No user used the help, although they were incentived during the training to use it if they had some doubt.

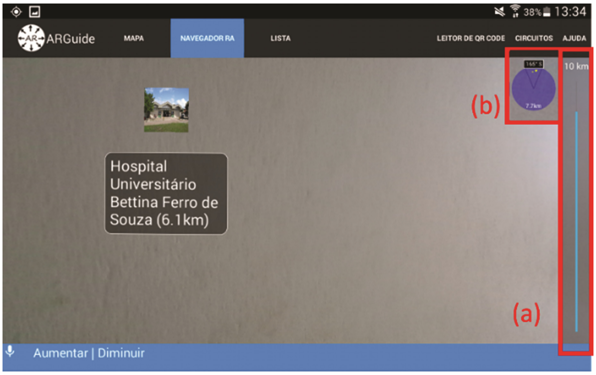

Fig. 6. AR browser view highlighting interface components

The units task 2c, 3c and 4c are the same (Select a POI), were the best received by the user. POI selection is a simple interaction, knowing the name of the POI the user can user it without the need to focus some marker in the screen, using the voice command as a shortcut. Two users highlighted speed reaction and visual feedback given by the application in POI selection. The suggestions of improvement for voice interaction can be listed as: Add more words for some commands, like the navigation ones.

- Give fail and success messages, with suggestions like: “Wanted to say:…?” in case of error.
- Clearly indicate what view between QR Code Scanner and AR Browser the user is using.

- Make the tab separator more visible and that tabs more separated from ARGuide logo.
- A customization option to select some commands before the user starts the usage of application.
- Redefine AR Browser layout. Figure 6(a) show the widget for manipulation of radius range and 6(b) the navigation radar, where each yellow dot is a POI of selected circuit. These points need more focus; most errors related to GUI could be avoided if these two widgets were clearer to the user.

6 Conclusion

Voice interaction allows a different user experience of an application. The MAR allows new types of exploration of a new place, or a new mode to explore old places. This work presented that is possible to unify this forms of interaction not overloading the user interaction. Efficient aspects of interaction was found, like the POI selection, as well as others can be improved, like the RA interaction. The results presented that a voice interaction in an RAM application are possible.

The selection of POI was the best solution to the interaction through voice interaction, therefore shortcuts interactions are feasible through voice interaction, (e.g. the user can speak the name of the POI if it is already known for the user, without search it in a vision). The RA browser's interface has some troubles in the voice interaction, and it will be redesigned next interaction. Based on the analysis of the results, a list of improvements was suggested. Among this improvements are the redesign of the RA browser' interface; to use more words in some commands; a custom interface to each user and error messages.

We expect that in the next application's iteration the improvements proposed in the session 5 may be made and a new test conducted, with a touch hybrid approach and voice to produce voice interactions guidelines to RAM applications.

References

1. Martínez, H., Skournetou, D., Hyppola, J., Laukkanen, S., Heikkila, A.: Drivers and bottlenecks in the adoption of augmented reality applications. J. Multimedia Theory Appl. **1**(1), 27–44 (2014)
2. Shneiderman, B.: Designing the User Interface: Strategies for Effective Human-Computer Interaction, 3rd edn, pp. 293–295. Addison Wesley, Reading (1998)
3. Vidakis, N., Syntychakis, M., Triantafyllidis, G., Akoumianakis, D.: Multimodal natural user interaction for multiple applications: the gesture – voice example. In: International Conference on Telecommunications and Multimedia – TEMU (2012)
4. Xia, L., Kai, K., Xiaochun, W., Dan, W.: Research and Design of the "Voice-Touch-Vision" Multimodal Integrated Voice Interaction in the Mobile Phone (2010)
5. Lee, K.B.; Grice, R.A.: The design and development of user interfaces for voice application in mobile devices. In: 2006 IEEE International Professional Communication Conference, pp. 308–320, 23–25 Outubro 2006

6. Teixeira, A., et al.: Speech-centric multimodal interaction for easy-toaccess online services – A personal life assistant for the elderly. In: 5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, DSAI 2013. Modeling and Simulation Design. AK Peters Ltd., Natick, MA (2013)
7. B.A. Delail, Weruaga, L., Jamal Zemerly, M.: CAViAR: Context aware visual indoor augmented reality for a university campus. In: Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology, (WI-IAT 2012), vol. 3, pp. 286–290. IEEE Computer Society, Washington, DC (2012)
8. Srinivasa, K.G., Jagannath, S., Akash Nidhi, P.S., Tejesh, S., Santhosh, K.: Augmented reality application: cloud based augmented reality android application to "know your world better". In: Proceedings of the 6th IBM Collaborative Academia Research Exchange Conference (I-CARE) on I-CARE 2014. ACM, New York (2014). Article 15
9. Markov-Vetter, D., Staadt, O.: A pilot study for augmented reality supported procedure guidance to operate payload racks on-board the international space station. In: IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 1–6, 1–4 October 2013
10. Doswell, J.T.: Augmented learning: context-aware mobile augmented reality architecture for learning. In: Sixth International Conference on Advanced Learning Technologies, pp. 1182–1183, 5–7 July 2006
11. Nielsen, J.: Usability Inspection Methods. In: Heuristic Evaluation. Katherine Schowalter, New York (1994)
12. Pyssysalo, T., Repo, T., Turunen, T., Lankila, T., Röning, J.: CyPhone—bringing augmented reality to next generation mobile phones. In: Proceedings of DARE 2000 on Designing augmented reality environments (DARE 2000), pp. 11–21. ACM, New York (2000)
13. de la Nube Aguirre Brito, C.: Augmented reality applied in tourism mobile applications. In: 2015 Second International Conference on eDemocracy & eGovernment (ICEDEG), pp. 120–125, 8–10 April 2015
14. Park, D., Nam, T.-J., Shi, C.-K.: Designing an immersive tour experience system for cultural tour sites. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems (CHI EA 2006), pp. 1193–1198. ACM, New York (2006). doi:http://dx.doi.org/10.1145/1125451.1125675
15. Feng, D., Meng, D., Zhang, Y., Weng, D.: Time machine: a mobile augmented reality system for tourism based on coded-aperture camera. In: 2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 10th International Conference on Autonomic and Trusted Computing (UIC/ATC), pp. 502–506, 18–21 December 2013