Narrative-driven Multimedia Tagging and Retrieval: Investigating Design and Practice for Speech-based Mobile Applications

Abhigyan Singh[1]*, Martha Larson*[2]

[1] Industrial Design, Delft University of Technology, Netherlands
[2] Electrical Engineering, Math and Computer Science, Delft University of Technology, Netherlands
a.singh@tudelft.nl, m.a.larson@tudelft.nl

Abstract

This paper presents a design concept for speech-based mobile applications that is based on the use of a narrative storyline. Its main contribution is to introduce the idea of conceptualizing speech-based mobile multimedia tagging and retrieval applications as a story that develops via interaction of the user with characters representing elements of the system. The aim of this paper is to encourage and support the research community to further explore and develop this concept into mature systems that allow for the accumulation and access of large quantities of speech-annotated images. We provide two resources intended to facilitate such work: First, we describe two applications, together referred as the 'Verbals Mobile System', that we have developed on the basis of this design concept, and implemented on Android platform 2.2 (API level 8) using Google's Speech Recognition service, Text-to-Speech Engine and Flickr API. The code for these applications has been made publically available to encourage further extension. Second, we distill our practical findings into a discussion of technology limitations and guidelines for the design of speech-based mobile applications, in an effort to support researchers to build on our work, while avoiding known pitfalls.

Index Terms: Mobile Speech Application, Multimedia Service, Narrative-driven design, Image Retrieval, Image Tagging, Android, Flickr

1. Introduction

Speech-based mobile applications, and specifically those that make use of a combination of speech and multimedia, are evolving rapidly. The data that such applications generate represent an interesting challenge for the research area of speech, language and audio in multimedia. However, speech has not yet established itself as a mainstream annotation modality for users who capture, save and share multimedia with their mobile phones. Instead, if users tag photos, they generally rely on text-based input forms. Before speech processing technology becomes truly relevant in the area of mobile multimedia, applications that allow users to use speech to tag and retrieve photos must establish their foothold. As long as use of such applications is not yet widespread, researchers will lack the critical mass of speech-annotated multimedia data that is necessary in order to address the question of how speech processing technology may support mobile speech-based image tagging and retrieval. This paper is motivated by the need for applications that break with existing conventions and practices in order to allow speech processing technology to realize its full potential to support users in capturing, sharing, viewing, and retrieving multimedia using mobile devices.

Upon first consideration, it appears nearly trivial to bring a photo-sharing website such as Flickr to a mobile phone by exploiting speech technology and using spoken tags and spoken queries to replace text-based user interaction. The fact that the telephone was originally developed as a device to support spoken communication, suggests, a priori, that voice input would be widely accepted by users as a mode for interacting with an image retrieval system. However, the idea that replacing text-tags with speech-tags is a simple switch soon reveals itself to be an overhasty assumption. Despite the research work that has been devoted to the topic of using speech to tag images, for example [1], [2], [3], there does not exist a single mobile application enabling speech-based tagging and retrieval of multimedia that has managed to enter mainstream use. In this paper, we take the standpoint that the challenges faced by such applications are not exclusively technical in nature, but rather have a critical dependency on the expectations and needs of users. The goal of this paper is to introduce a new design concept for mobile speech-based multimedia tagging and retrieval applications that tackles the challenges of handling the interaction between technological limitations and how users expect the system to work and what they want to achieve.

The key insight of our design concept is that the interaction of the user with the tagging and retrieval application can be conceptualized as a narrative. Within the storyline of this narrative different characters interact. These characters correspond to different elements of the system, and have personalities that reflect the function, speed and reliability of these elements. If the storyline is designed so that it is both engaging and easy to understand, the user will be able to formulate a useful model to understand the inner mechanics of the system that will promote acceptance and patience.

The rest of the paper is organized as follows. In Section 2, we discuss the preliminary investigation that led us to be concerned with the interaction of user expectations and technological limitations and to arrive at the concept of narrative-driven multimedia tagging and retrieval. In Section 3, we present background information on narrative structure and also our narrative-based design concept. Then, in Section 4, we describe our experiences in applying this design concept in the development of a particular mobile application for speech-based tagging and retrieval of images, called the 'Verbals Mobile System' (where 'verbal' refers to a voice tag). Finally, in Section 5, we distill our experiences into a series of lessons learned that will inform the development of other speech-based mobile applications for multimedia that apply the same design concept. We finish with a conclusion and outlook onto future work.

Proceedings of the First Workshop on Speech, Language and Audio in Multimedia (SLAM), Marseille, France, August 22-23, 2013.

2. Preliminary Investigation

The troublesomeness of speech recognition errors for systems that allow users to speech-tag their photos has been pointed out in the literature [2], [4]. The focus of our investigations was set specifically on speech-tagging in mobile environments, and we began with an informal study of existing mobile applications and a field study. During the field study, we tested mobile speech recognition technology in a series of real-world environments. Our investigation identified a set of key design challenges for the design and implementation of mobile 'voice' interfaces that connect with a remote image sharing server. The first set of challenges is technical in nature: Speech recognition in mobile environments (i.e., settings characterized by every-day noise conditions) falls far short of the recognition levels of human listeners in the same environments. Mobile environments are more challenging than other environments in which speech recognition technology is used because they are uncontrolled and highly variable (including background babble, birds, traffic, music, construction, wind). The strength of the signal received by a smartphone can vary with environment (e.g., at high-altitude, in forests, or in basements), meaning that connection with a remote image-sharing server is not stable. Mobile speech recognition technology that requires an Internet connection also suffers from delay and disconnection. The second set of challenges is related to users: The use of speech input is restricted by social and cultural norms that apply in both public and private spaces. Speech is for this reason not always acceptable in certain civic settings (e.g., in certain parts of hospitals or during a lecture/meeting) and may also be more disruptive than text input when a user is simultaneously engaged in social interactions with other people. Misrecognitions and delays caused by the technical challenges mentioned above can be a source of enormous frustration for users, causing them to experience the application as tedious to use or not worth the effort.

A critical characteristic of these design challenges is that the technical challenges are not separated from the user challenges, but rather the two sets of challenges are interdependent. Based on this observation, we concluded speech-based mobile applications must arise from a convergence of design and technology. We then set for ourselves the objective of developing a design concept that would accommodate both the needs of users and the technical restrictions of mobile speech, focusing on two main questions: (1) How to improve the tolerance of users for mobile multimedia applications which are considered technically to be 'not-perfect' or 'still-evolving'? (2) How to deal with possible errors, introduced by automatic speech recognition into tags and queries, in a manner acceptable to users?

As the basis for our design concept, we chose to conceptualize user interaction with the mobile application as a narrative. The aim of the narrative is to engage users and to provide useful explanations for the system's unexpected behavior. In the following section, we describe the 'Verbals Mobile System', which we developed as an application of this design to the task of image tagging and retrieval.

3. Narrative-Driven Mobile Speech Design Concept

Narrative structure could be understood as consisting of two parts: Content and Form. While content consists of story, i.e., characters, events, and conflicts, the form deals with plot, i.e., how the story is told or narrated. A narrative adds to the meaning generation process and engages an audience by facilitating interpretation of the story [5], [5], [6], [7]. Narrative structures are widely utilized in variety of media like newspapers, television advertisements, documentaries, films, and games to communicate with the audience [6], [7].

Upon first consideration, there are a seemingly endless number of stories, making it difficult to decide which storyline would be most appropriate for use in a mobile speech-based application. Ideally, the story chosen should be appropriate for any possible user of the system. Closer examination of literary and anthropological research reveals that narratives generally fall into broad patterns that share wide appeal. The folklorists and anthropologists that have analyzed and discussed narrative patterns in folklores and myths include Tvzentan Todorov, Vladimir Propp, Claude Levi-Strass, Roland Barthes and Joseph Campbell. Vladimir Propp's analyzed Russian folktales and discussed 31 narrative functions in his seminal book, 'Morphology of Folk tales', first published in 1928. He defined narrative functions as smallest possible unit of a narrative and described a certain order of appearance narrative functions in Russian folktales [8]. Propp's work has influenced work of many theoreticians and practitioners including Joseph Campbell, who analyzed myths in various cultures in his book 'The Hero with a Thousand Faces' [9], [10]. Campbell describes a pattern called Hero's Journey, which has been used in many Hollywood movies including Star Wars. It is the Hero's Journey narrative pattern that we chose to build upon in our design concept because it involves a single protagonist (who is represented in our concept by the user), overcoming unexpected obstacles (which are represented by unpredictable behavior of system elements such as speech recognition) in order to reach a goal (tagging or retrieval of images).

Various mediums offer varying benefits and limitations for incorporating narrative structures. The narratives could be adapted to these strengths and limitations [6], [7], [11]. For instance, a three-hours adventure film may have multiple subplots and numerous characters, whereas a thirty-second television commercial may have a single plot and a character. Narrative structure for mobile phones has not been explored much as yet. However, with their increasing ubiquity and technological advancement, mobile phones as a medium present an evolving and challenging platform for use of narrative structures [6], [7], [12]. Mobile phones facilitate various modes of interactions not possible in mediums like films or television. Mobile phones, like the World Wide Web itself, is an interactive medium rich with possibilities for multi-modal interactions (e.g., text, speech, touch) and the simultaneous use of multimodal features (e.g., audio, video, images). Further, the mobile phone is usually seen as a personalized object primarily used by a single user. This provides possibilities for a much more intimate interaction with the user. Due to these characteristics, mobile phones offer a rich platform for exploring narrative structures.

Our design concept enhances the tolerance of users for imperfect mobile phone technology by using a narrative in order to provide an explanation for unexpected behavior, such as disconnections, delays, and speech recognition errors. Because they are engaged with the narrative, users will also experience the time needed for processing and transmission to be less frustrating. In the next section, we apply our design concept to develop the 'Verbals Mobile System'.

4. Verbals Mobile System

This section first describes the narrative structure that was developed for the 'Verbals Mobile System', which we developed as an application of this design to the task of image tagging and retrieval.

4.1. Narrative Structure

As the basis for our design concept, we chose to conceptualize user interaction with the mobile application as a narrative, which we create by overlaying a dialogue with characters and a storyline. As previously mentioned, we build on the narrative pattern of 'Hero's Journey'. We brainstormed on various possibilities for a story and a plot and narrowed them down to the theme of travelling back in time in combination with a common human interest to view images of events in the past. As a result, the narrative of 'Verbals Mobile System' is based on the theme of 'Communication with the Past' and the genre of 'adventure'.

The narrative emerges via the interaction of a set of characters who pursue individual goals and each have their own personalities. Different realizations of the interaction are conceivable. For example, it would be possible for all characters to be explicitly instantiated by animation or voices seen and heard by the user and to interact with the user directly. We choose for a realization in which the user interacts with only a single, central character and that the other characters are not directly represented. Rather their actions are reported to the user by the central character. This choice simplifies the user interaction with the system and allows us to map the narrative onto a dialogue between two persons, the user, who is the protagonist of the narrative and the central character, who is directed by the user.

In the remainder of this section, we discuss the characters in our narrative (summarized in Table 1) and the mechanics of the plot. Note that the narrative-based designed concept that we proposed is not restricted to use of these characters or mechanisms. Rather, the specifics that we present here serve as an example of how the cast of characters can be established, and what the correspondence should be between the characters and the elements of the system.

Central character: 'Pica' is the central character and is a bird that can be directed by the users of our applications. We selected a bird since the human users of the applications could naturally associate birds with flying and with the activity of sending messages or communication with distant lands. The type of bird is a European Magpie, which was chosen because it is a common bird in the Netherlands and is also well known to for its habit of carrying shiny objects off and for its mischievous intelligence. We characterized 'Pica' as having a special ability to fly to 'The Past' and some ability to understand human voice. A user interacts with Pica and can direct her to travel back in time to access the human memories in the 'Past'. When users start the 'Verbals Push' application, they hear the voice of Pica saying, 'I am Pica, the magpie, I can carry your tags and images to the invisible Land of the Past. Select an images that you wish to send to the Garden of Human memories in the Past'.

Mentor: 'Google', is characterized as the omnipresent mentor of Pica, who helps her in understanding and interpreting human voice. A user can speak to Pica and send her on a journey to the 'Past'. The user can direct Pica to bring images from 'Past' and can also send images and some tags to 'Past'. The 'Past' appears in the narrative as a distant vast land holding shared human memories. Users all live in the 'Present' and any moment before the 'Present' is in the 'Past'. Please see Table-1 for the list of characters and their roles in the Verbals Mobile System.

Table 1. *List of characters and their roles.*

Character	Role in narrative	Role in the application
Pica [Protagonist]	Pica, a female magpie has the ability to fly back in time to reach 'The Past'. Pica could carry images and tags	Interaction with the user
Google [Mentor]	An omnipresent mentor, who helps Pica in understanding human voice	Accessing Google speech recognition service and obtaining results
Flickr [Guardian]	The guardian of an invisible valley in the distant clouds having an entrance to 'The Past'. To enter 'The Past' for the first time Pica has to provide a secret code	Fetch/Post images and tags from Flickr, Flickr account authentication
Zaat [Antagonist]	A wind demon in 'The Past' who shoots fast vertical winds to slow-down, disorient or snatch images from Pica in her journey	To deal with delay or error in image fetch/post
Bazooka [Antagonist]	A sonic demon that fires rapid noise beams to disrupt Pica's communication with Google. Bazooka does not like silence and human voice	To deal with incorrect speech recognition of a user's voice input

Quest: To enter the 'Past', Pica has to provide secret code to 'Flickr'. 'Flickr' is the guardian of an invisible valley in the distant clouds having an entrance to the 'Past'.

Conflict and challenges: Bazooka, a sonic demon does not like silence and human voice. Bazooka tries to attack with rapid noise beams whenever it gets to know of a communication between a human (user) and Pica. The noise beams could disrupt Pica's communication with 'Google' and hence lead to misinterpretations. So, the first challenge for a user is to successfully communicate with Pica so that Bazooka does not get to break the communication. Zaat, a wind demon in the 'Past' shoots fast vertical winds to slow-down, disorient or snatch images from Pica in her journey. So, the second challenge for a user is to succeed in either sending images and some tags to the 'Past' or retrieving images from the 'Past'.

Reward: Users who succeed in dealing with the conflicts and the challenges get to see ten images from the 'Past' as identified by their spoken tags (Verbals Pull) or their selected image and spoken tags successfully stored in 'The Past' (Verbals Push).

4.2. Architecture of the Verbals Mobile System

In this section, we discuss the Verbals Mobile System's architecture. The system consists of two separate applications: one for speech tagging (referred to as 'Verbals Push') and for speech search (referred to as 'Verbals Pull'). We have implemented the architecture as an Android phone application using Google Speech API, and Flickr API. The open source source-code of the Verbals Mobile System, including both the Verbal Push Application[1] and the Verbal Pull Application[2], is available on line. We have also made a demo video[3] available.

Verbals Push Application: The architecture of the Verbals Push application for speech tagging is depicted in Figure 1. The dialogue manager is the heart of the architecture. The user interacts (through the mobile phone's user interface) with the dialogue manager and the dialogue manager based on the current state of the application decides on subsequent actions and provides user feedbacks. For instance, when a user starts the application the dialogue manager introduces the application (including the character of 'Pica' as part of the narrative) and facilitates the user to select an image from phone's memory. Further, the dialogue manager uses Google Speech API to identify user's spoken-tags (or 'verbals'). Google Speech API sends user spoken-input to Google Speech servers and receives results. The speech recognition results are shown on phone's interface and dialogue manager requests for an implicit confirmation from the user. On receiving a 'go-ahead' from the user, dialogue manager connects with Flickr API, sends user authentication information and then pushes the selected image and the spoken-tag to the users Flickr account. Now, the user should see the image and the tags on his/her Flickr photostream. The dialogue manager uses the phone's text-to-speech engine to provide spoken feedback to the user and facilitates dialogue delivery for the narrative. The dialogues are implemented in an XML file format.

Verbals Pull Application: The architecture of the Verbals Pull application for speech-based image retrieval is depicted in Figure 2. The dialogue manager is, again, the heart of the architecture. The user interacts (through the mobile phone's user interface) with the dialogue manager and the dialogue manager based on the current state of the application decides on subsequent actions and provides user feedbacks. For instance, when a user starts the application the dialogue manager introduces the application (including the character of 'Pica' as part of the narrative) and encourages the user to speak a location. Further, the dialogue manager uses Google Speech API to identify user's spoken-tag (in this case a location name). Google Speech API sends user spoken-input to Google Speech servers and receives results. The speech recognition results are shown on phone's interface and dialogue manager requests for an implicit confirmation from the user. On receiving a 'go-ahead' from the user, dialogue manager connects with Flickr API, and then requests for ten random public images on Flickr that are tagged with the user's spoken-tag (a location name). The Flickr API returns the images and the images are shown on the phone's user interface. Again here the dialogue manger uses phone's text-to-speech engine to provide spoken feedback to the user and facilitates dialogue delivery as part of the narrative. The dialogues are implemented in an XML file format.

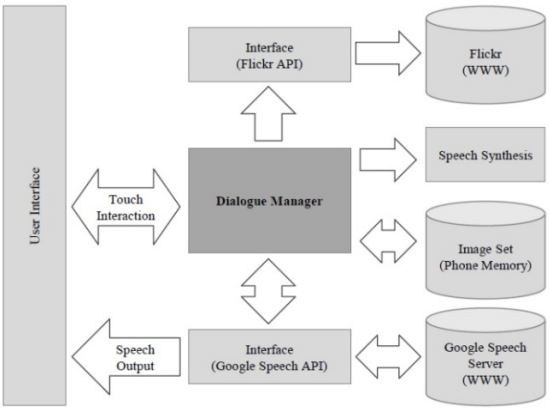

Figure 1: *Verbals Push Application.*

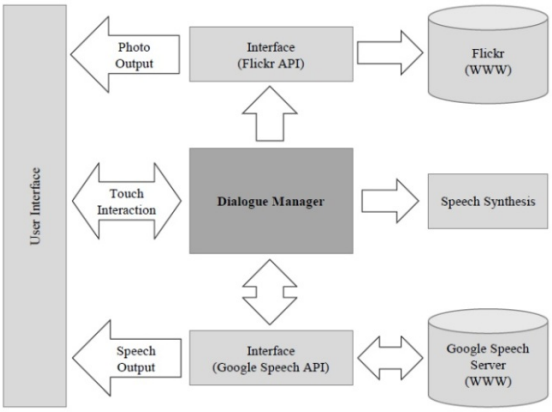

Figure 2: *Verbals Pull Application.*

5. Mobile speech in practice

In this section, we briefly present our general observations, first covering technical and then design aspects.

5.1. Technical Challenges and Limitations

Android platform: Before starting development it is important to realize that the Android platform based speech application development will require a real mobile phone in the development process as the Android emulator does not cater to voice input.

Dialogue manager: Implementing a dialogue manager is the key aspect for a narrative driven mobile speech applications. Android (2.2) framework requires a low-level implementation of dialogue manager. Many conventional server-side dialogue-driven applications like IVR systems use VoiceXML, which structures and simplifies the implementation of dialogues manager by facilitating a high-level implementation. However, VoiceXML (version 2.0 and 2.1) is primarily designed for server-side implementations and not suitable for client-side implementations as in case of Verbal's Mobile System. The upcoming version of VoiceXML may address this issue.

Mobile speech recognition: The Google Speech API, as of the time of writing, offers limited functionalities, a factor that

[1] https://github.com/abhigyan/Verbals-Push
[2] https://github.com/abhigyan/Verbals-Pull
[3] http://youtu.be/rnXtBsVgEII

constrains the features and interaction models that can be implemented in mobile speech application by application developers. For instance, it is not possible to record the user's voice input while using the Speech Recognition service or to retrieve the recorded audio-file from speech servers. A user click is needed to start the Speech Recognition as the mobile speech recognition cannot be in 'always on' mode. For this reason, design for speech-enabled application on mobile phone needs to balance 'speech' and 'touch' input. Google Speech Recognition's language options are evolving and at the time of writing there are limitations in vocabulary. For instance, during the time of implementation of Verbals Mobile System, Dutch Language and English (Dutch) Locale were not available. Implementing and designing mobile speech using Android (API 2.2) requires an Internet connection as the user's speech input is sent and results are received from Google's speech servers. This dependence on communication with speech servers brings possibilities of delays, errors and disconnections.

Mobile speech synthesis: The Android's default Text-to-Speech Engine (PICO) provides limited options. For instance, for speech synthesis only a 'female' voice option is available. Similarly, various speech synthesis audio-effects like echo, tempo, and accents are not yet available. This limits the use of narrative based on multiple characters as such a scenario would require variation in dialogue delivery in terms of pace, voice, pitch to depict contrast, variation in moods and personalities of the characters.

Multimedia sharing platform: Certain functionalities (like Image Post) using Flickr API requires user authentication. At the time of implementation of the Verbals Mobile System, the Flickr API was migrating to OAuth authentication scheme. Although this new mechanism is enhanced and more secure compared to the earlier version of authentication, it requires investing additional implementation time. Time and care is needed in selecting an Android-Java Flickr library. There are varieties of options available but many libraries offer incomplete functionalities and some are not well documented.

5.2. Guidelines for Mobile Speech Applications

Our guidelines for mobile speech application design emphasize user aspects, since we anticipate that specific aspects of the needs of users will endure, even after technical limitations have been addressed. First, we note that users needs in a certain sense run ahead of technology. Even as mobile phone coverage continues to improve globally, users will continue to find new places to capture and share multimedia (e.g., underground, under water). Multimedia content contains increasingly more information, for example, we can anticipate a move from images to video to 3D video. It is important not to assume that bandwidth is cheap, and its price might be the limitation on use for some users.

Second, we note that narrative-driven dialogue design has an enormous potential to engage users with speech based systems and improve their experience. However, a variety of factors shape the aesthetics of dialogues and narrative structure for that are appropriate for mobile platforms. Long and complex dialogues that maybe fine for interactive applications on the Web or for films, but these could create significant problems and reduce user experience when used in mobile speech applications. The aesthetics of vocal delivery is important: 'machine-like', 'young adult', 'feminine' and having 'native English accent' can all contribute to building the narrative. However, these can also have an impact on users' individual emotional response and must be taken into consideration. In order to target broad appeal, various popular genres of narrative like humor, horror or mystery could be leveraged.

6. Conclusions and Outlook

We have presented a narrative-based design that we have developed in order to reconcile technological limitations of mobile speech-based applications with user expectations. We start from the observation that the key to wider uptake of mobile speech-based applications for multimedia tagging and retrieval is not exclusively technical in nature. Rather, such a solution requires careful consideration of what users expect applications to do, and what they actually can do, given technological limitations.

We have presented an example narrative that builds on the 'Hero's Journey' narrative pattern. The individual characters involved in the narrative correspond to individual elements of the application, and are given personalities that reflect the unexpected behavior of these elements, so that they can serve to 'explain' to users why speech-based tagging and retrieval does not always precede along a smooth and predictable path, but rather encounters technical limitations.

We would like to note that we anticipate our approach will remain relevant, even as speech-based mobile technology continues to develop. A notion of "error free" speech recognition is difficult to formulate, since a speech recognition system that never makes an error must recognize speech better than a human being. However, even should "error free" speech recognition measured by any reasonable notion be achieved, tagging and retrieval systems still stand to benefit from collecting more and richer data from the users, in terms tags and queries that are more specific. Here, the engagement of a narrative-based system could help to extend users' patience and direct their formulation of tags and queries.

We would like to point out that smartphones offer possibilities for multi-modal interactions for narrative structure, going far beyond what we have covered here. Moving forward, we see opportunities for narrative-driven speech applications to be enhanced by integrating rich user context information (e.g., geo-location) and sensor information (e.g., from accelerometers) as narrative elements. We also are interested in the question of speech-based annotation and retrieval of video. The fact that video is a temporally continuous medium opens new challenges e.g., will users associate speech annotation with particular parts of the video, or with the whole video? The fact that videos may already contain speech might make it undesirable to add a second layer of spoken annotation. For example, users might be able to easily "speak over" existing speech to add or listen to annotations.

The number and variety of challenges left open by our work provides a rich field for future investigation, which must include design and implementation of mature mobile speech-based tagging and retrieval applications as well as testing and refinement. We hope that with this work, we have opened the door for such a future and have provided tools and resources that might prove useful to support it.

7. Acknowledgments

We thank the STITPRO Foundation http://www.stitpro.nl/ for their generous funding that supported this work.

8. References

[1] Chen, Jiayi, Tan, Tele, Mulhem, P., and Kankanhalli M. "An Improved Method for Image Retrieval Using Speech Annotation." In Proceedings of the International Conference on Multi-Media Modelling, 12-30, 2003.

[2] Hazen, T.J. Sherry, B. and Adler, M. "Speech-Based Annotation and Retrieval of Digital Photographs," In Proceedings of Interspeech 2007, 2165-2168, 2007.

[3] Kalashnikov, D.V., Mehrotra, S., Jie Xu, and Venkatasubramanian, N., "A Semantics-Based Approach for Speech Annotation of Images," IEEE Transactions on Knowledge and Data Engineering, 23(9):1373-1387, 2011.

[4] Rodden, K. and Wood, K.R. "How do people manage their digital photographs?" In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems *(CHI '03).* ACM, New York, NY, USA, 409-416, 2003.

[5] Wilson, K., "Narrative", Online: http://mediaknowall.com/as_alevel/alevkeyconcepts/alevelkeycon.php?pageID=narrative, accessed on 1 July 2013.Block, B., "The Visual Story: Creating the Visual Structure of Film, TV and Digital Media", Second edition, Focal Press, Oxford, UK, 2007.

[6] Miller, C. H., "Digital Storytelling: A creator's guide to interactive entertainment", Second edition, Focal Press, 2008.

[7] Alexander, B., "The New Digital Storytelling: Creating Narratives with New Media", Praeger, California, 2011.

[8] Propp, V., "Morphology of the Folktale", L. Scott [Translator], Second edition, University of Texas Press, Austin, Texas, 1968.

[9] Wilson, K., "Propp's Analysis of Folk Tales", Online: http://www.mediaknowall.com/as_alevel/alevkeyconcepts/alevelkeycon.php?pageID=propp, accessed on 1 July 2013.

[10] Campbell, J., "The Hero with a Thousand Faces", Third edition New World Library, California, 2008.

[11] Lindley, C. A., "Story and Narrative Structures in Computer Games", in B. Bushoff [Ed], Developing Interactive Narrative Content, München: High Text Verlag, 2005.

[12] Epstein, M. and Vergani, S., "History Unwired: Mobile Narrative in Historic Cities", AVI '06, Venezia, Italy, 302-305, 2006.