PHOTOPAL: Companionship, Sharing and the Digital Echo

Oli Mival, Brian O'Keefe, Jay Bradley, Nena Roa Seiler & David Benyon
Centre for Interaction Design
Napier University, Edinburgh, EH10 5DT
o.mival@napier.ac.uk - b.okeefe@napier.ac.uk
n.roa-seiler@napier.ac.uk - j.bradley@napier.ac.uk - d.benyon@napier.ac.uk

ABSTRACT
This short paper introduces the research of the 14 partner, EU Framework 6 project, COMPANIONS. It focuses on the development of PhotoPal, a multimodal system harnessing the cutting edge of speech recognition, natural language understanding and dialogue management, alongside advances in touch, multi-touch and gesture driven interactions. PhotoPal allows for a far more natural engagement with a user's digital photo collection, the tasks associated with it and the tools enabling those tasks to be achieved. Removing the mediation of a keyboard and mouse paradigm and introducing a speech, touch and gesture driven interaction experience enables the intuitive familiarity of dealing with analogue photographs whilst retaining the huge storage, metadata addition, editing, sorting and distribution advantage of the digital format.

Utilising the power of natural language understanding has the considerable addition of enabling a user to establish not only a hugely rich and sophisticated pool of metadata, but also a very personal and unique audio narrative. The combination of this image specific metadata (such as content, time, location) alongside more abstract elements (such as why a photo was taken or what emotion it provokes) gives rise to the potential of establishing rich life narratives, or Digital Echo.

Furthermore, the conversational nature of the interaction experience positively encourages and facilitates a social experience, both with the PhotoPal system itself (and it's representative avatar) and with family, friends and beyond, both locally and remotely. This experience not only goes beyond locale, but also time and potentially life time.

Author Keywords
Companionship, Photography, Multimodality, Internet, Avatar, Natural Language Understanding, HCI.

INTRODUCING THE COMPANIONS PROJECT
The COMPANIONS project is a 12 million euro, 4-year EU Framework Programme 6 project, involving a consortium of 14 global partners across 8 countries. It's aim is to develop a personalized conversational interface, one that knows and understands its owner, and can act as an alternative access point to resources on the Internet, all the while nurturing an emotional involvement from it's owner/user to invoke the shift from interaction to relationship. On a technical level it intends to push the state of the art in machine based natural language understanding, knowledge structures, speech recognition and text to speech. With these technical developments will come advanced interaction design elements.

The COMPANIONS design studio is based at Napier University's Centre for Interaction Design and is primarily responsible for the overall interaction design and the aesthetic development of COMPANIONS interfaces.

COMPANIONS will learn about their owners: their habits, their needs and their life memories. This will allow them to assist with carrying out specific Internet tasks, which will be facilitated by having complex models of their owners, by which we mean whole-life-memories, or coherent autobiographies, built from texts, conversations, images and videos. Some of this will already be in digital form, but some will be information gleaned from conversations with the COMPANION.

COMPANIONS will be autonomous and have original aspects of persistent human personality to establish loyalty and trust between users and such agents. They will be sensitive to limited emotion in speech and to the content of images, and will be themselves capable of demonstrating emotional/affective behavior through speech and visual appearance (e.g. an avatar on a PC screen or mobile phone). COMPANIONS will also, by communicating with each other enable and enhance communication between the human users, rather than only between humans and these machine artifacts.

CHARACTERISTICS OF COMPANIONS
There are clearly many challenges for 'companion technology'. Companions are a development of agents. Agents appear in the literature as software agents, interface

agents or embodied conversational agents (ECA). ECAs have typically been more concerned with behaviors (Pelachaud, 2005). Interface agents have focused on dealing with some specific aspects of HCI. Some early thoughts on interacting with interface agents did highlight speech as a key element (Norman, 1994). In software the traditional model of agents is that they have beliefs, desires and intentions, sometimes referred to as BDI agents. Companions draw upon all of these, and on spoken natural language technologies. It is this combination, which we believe will shift interactions into relationships.

Bickmore and Picard (2005) argue that maintaining relationships involves managing expectations, attitudes and intentions. They emphasize that relationships are long-term built up over time through many interactions. Relationships are fundamentally social and emotional, persistent and personalized. Citing Kelley they say that relationships demonstrate interdependence between two parties – a change in one results in a change to the other. Relationships demonstrate unique patterns of interaction for a particular dyad, a sense of 'reliable alliance'.

It is these characteristics of relationships as rich and extended forms of affective and social interaction that the project is trying to tease apart so that as to provide advice for people designing companions. Digesting all experiences to date we describe companions by looking at the characteristics of companions in terms of utility, form, personality, emotion, social aspects and trust.

EXPLORING A PHOTO COMPANION

How a photo is taken is no longer the issue (digital imagery comes from sources never originated as "cameras", eg mobile phones, screenshots etc), rather their organization, distribution, styling and review comes to the fore.

Yorik Wilks, who initiated and leads the COMPANIONS project, proposed the idea of a companion to help older people sort out their photos and life memories. With many people now having thousands of digital photos, sorting them, classifying them and organizing them becomes a huge issue. This raises the question of how the average person with no (or limited) classification or editing skills could even begin to make a coherent shape of such a mass of data. Despite the emergence of a multitude of on and offline photo logging, collation, editing and distribution applications (from Flickr.com and Picassa.com to Photoshop and iPhoto), the fact remains that for the majority, digital photos still lack the emotional significance of traditional analogue collections and shared social experiences (Kirk, et al., 2006). It is with these issues in mind that the concept of the PhotoPal has been examined.

Starting in January 2007, a series of Wizard of Oz (WoZ) studies were undertaken to explore peoples attitudes to discussing their photographs with an apparently intelligent computer system, represented by a variety of avatars, voices and interfaces, see Figure1 for two examples.

Figure1.
Screenshots of two iterations of the user interface for the PhotoPal WoZ (companion is represented by avatar in top example and by speech alone in bottom example)

As of October 2007, 65 sessions involving 55 participants had been recorded. The dialogue between the PhotoPal system companion (played by a remotely based human wizard controlling the audio visual output of the interface) and the participant functions not only as a qualitative source of data for the investigation of design requirements, but is also used to inform dialogue management, knowledge structure, speech recognition and natural language understanding systems of specifics within the photo domain.

The PhotoPal concept can be considered a digital photo editing, sorting and sharing companion. The owner interacts through natural language dialogue with the companion, represented by an on screen avatar. Through the process of talking about the quality of their photos ("that's a little dark"), the location where they were taken ("oh, I took this picture, this in my garden"), the time it was taken ("this was on my birthday") and the content ("on the right is my brother, he's holding his daughter Julie in his arms"), the companion is able to fix quality issues, organize and display folders by content location and date, and most importantly develop a rich amount of metadata. This interaction – where rich descriptions in natural language are used to identify the semantics and affective aspects of the photos is being called "Talk-to-Tag". Furthermore, the PhotoPal companion can then use the social and familial networking knowledge structure that the Talk-to-Tag process has generated to engage in smart sharing. For example, pictures from a family gathering can be sent to the interactive smart photo frames of the family members who were there, or perhaps

those that were not. Having the photos tagged in this way will also facilitate the owner reminiscing with the companion and hence allowing PhotoPal to gather even more information about the details and relationships depicted in the photographs.

The use of the WoZ methodology has also allowed for a unique design methodology in which various functionalities of the wizard are replaced as the technology reaches maturity, eg, face recognition. The project has also recently developed a series of demonstrators in which the wizard has been removed entirely, and all conversations and functionality are achieved through touch, gesture and the speech system alone. These demonstrators are limited in scope by parameters such as vocabulary, but are extremely useful as investigatory tools for examining multiple photo and multi user (both local and remote) interactions and experience.

PHOTOPAL FUNCTIONALITY AND THE DIGITAL ECHO

PhotoPal research is currently primarily focused on companionship for recently retired older users. We are concerned with life memories, reminiscing and how these seemingly fleeting activities can be captured through the combination of dialogue, image and artificial companionship. In other words, we are seeking to enable a scenario in which it is possible to transmit memories through the Senior Photo-Companion long after the owner is deceased.

There are multiple functionalities proposed for a Senior Photo-Companion. For example, PhotoPal could send a selected photo to an identified friend or relation, because PhotoPal can access the necessary addresses and functions to do this. PhotoPal would be able to discard blurred pictures, but would be unlikely to argue that one was a bit too dark (unless is was much too dark). This sort of judgment should rightly come from the human in this relationship. Leave PhotoPal to perform the function of lightening the picture, but leave the human to judge which pictures to lighten. This reflects the discussions on the allocation of authority and functions in HCI. The 'instrumental support' (Bickmore and Picard, 2005) provided by a companion is a key part of relationship building. Also, the user can place images to tell a story, as story telling is an important facet for our reminiscing requirements. These stories can be categorized by event, people or places, or simply as a chronological biography.

However, the Senior Photo-Companion is only one of many Photo Based Companions the COMPANIONS project have set out to explore and develop. For example, scenarios involving new families (ie, young couples with their first child) and people traveling or working abroad for long periods (eg in the armed forces, students on gap year) are being explored. Specifically these scenarios examine how the PhotoPal can facilitate the sorting, storing and sharing of new life narratives (new born baby), current narratives (those traveling) and retrospectives (older users and the recently retired) thus creating a *Digital Echo* of their lives.

CONCLUSION

Combining natural speech with touch and gesture removes the traditional mediation of a keyboard and mouse interface. The impact of progressively more sophisticated display technologies such as Microsoft's Surface and Apple's mulitouch iPhone and iPod Touch are testament to the intuitive nature of touchscreens as an interaction interface of choice in the photo domain. More importantly, removing a users Photos from the (largely conceptual) "shoebox on a hard drive", allows for far more exciting photo activities such as the smart metadata driven sorting and sharing of individual and collective memories.

PhotoPal is about facilitating the addition (both automatically and consciously) of that metadata invisibly, and utilizing it to visualize, conceptualise, sort, style and share a users photos in a way that is painless, obvious, helpful and useful. Moving away from visual search into conceptual (and literal) search is a paradigm shift facilitated by the application and successful implementation of metadata.

This not only provides opportunity for a new photography based companion but also gives the opportunity to move from human-computer interactions to human-companion relationships facilitating human-human relationships. People will stop being computer users and will become companion owners, and this in turn will bring with it a whole new social experience.

ACKNOWLEDGMENTS

This work is funded by the European Commission under contract IST 034434.

REFERENCES

1. Bickmore T. and Picard R. (2005) Establishing and maintaining long-term human-computer relationships. ACM Press.
2. Kirk, D. Sellen, A. Rother, C. and Wood, K. (2006) Collecting and editing photos: Understanding photowork Proceedings of the SIGCHI conference on Human Factors in computing systems CHI '06 761 – 770
3. Mival, O. and Benyon, D. R. (2007) Introducing the COMPANIONS project: Intelligent, persistent, personalised multimodal interfaces to the internet. In Proceedings of AISB2007, Newcastle
4. Norman, D. (1994) How might people interact with Agents? Communication of the ACM vol 37, N°7 July 68-71
5. Pelachaud, C. (2005) Brave new topics 2: affective multimodal human-computer interaction: Multimodal expressive embodied conversational agents Proceedings of the 13th annual ACM international conference on Multimedia MULTIMEDIA '05 683 – 689.