Digital Pens as Smart Objects in Multimodal Medical Application Frameworks

Markus Weber, Christian H. Schulz
German Research Center for AI
{firstname.surname}@dfki.de

Daniel Sonntag, Takumi Toyama
German Research Center for AI
{firstname.surname}@dfki.de

ABSTRACT
In this paper, we present a novel mobile interaction system which combines a pen-based interface with a head-mounted display (HMD) for clinical radiology reports in the field of mammography. We consider a digital pen as an anthropocentric smart object, one that allows for a physical, tangible and embodied interaction to enhance data input in a mobile on-body HMD environment. Our system provides an intuitive way for a radiologist to write a structured report with a special pen on normal paper and receive real-time feedback using HMD technology. We will focus on the combination of new interaction possibilities with smart digital pens in this multimodal scenario due to a new real-time visualisation possibility.

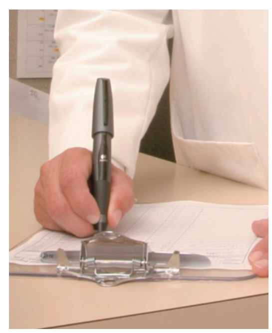
ACM Classification Keywords
H.5.2 User Interfaces: Input Devices and Strategies, Graphical HCIs, Prototyping

Author Keywords
Augmented Reality, Medical Healthcare, Real-time Interaction

INTRODUCTION
A standard reaction of computer-affine people to this question is that they are much faster with a keyboard. And it is true, handwriting as an input metaphor is a very slow interaction process when you are trying to input information into the computer, especially for those people who learned typing rapidly with the keyboard.

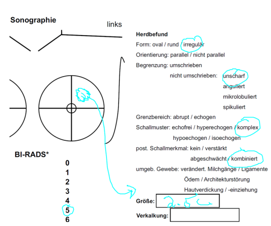
However, people still use a pen for putting down information on paper and a lot of processes are still based on paper documents. In radiology practices paper reporting have been established over the last 20 years. However, this situation is not optimal in the digital world of database patient records. Digital records have many advantages over current filling systems when it comes to search and navigation in complete patient repositories called radiology information systems. In fact, modern hospital processes require digital patient reports. The current practice in hospitals is that dictated or written patient reports are transcribed by hospital staff and sent back to the radiologist for approval. The turn-over time is 2-30 hours and the process inefficient and also prone to error.

THE MOTIVATION
In [4], we presented an interaction method that shows how a radiologist can use our special mammography paper writing system to conduct a full mammography patient finding process. This digital pen based interface enables radiologists to create high-quality patient reports more efficiently and in parallel to their patient examination task. Thereby, he or she uses a digital pen-based interface where the user can write on normal paper which is printed with a light-grey dot pattern in order to allow for the recognition of the writing. Nevertheless, this solution still requires a workstation to provide feedback of the real-time recognition results; hence it limits the mobility of the pen-paper approach. Moreover the doctor has to constantly change his or her sight from the paper to the screen and back, thus the doctor cannot focus on the patient.

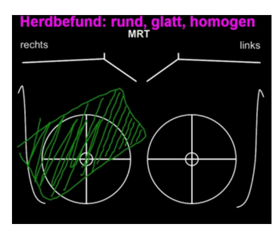


(a) Radiologist using digital pen to diagnose patients.

(b) Brother's AiRScouter HMD.



(c) Annotated form with region of interest marking the selection of associated anatomical concepts.



(d) Visualisation of recognised results in the HMD display.

Figure 1: Mobile pen-based system.

In [5] we have developed a mobile working station for the radiologist that is called RadSpeech. In the following we have extended the system for the use of multiple mobile stations combined with one stationary screen installation (see Figure 3a). The motivation of the setting here, was to provide hands-free interaction with a medical system using the natural language. Basically, the scenario describes how the medical expert retrieves medical images of one specific patient and then continues to make his finding by attaching semantic information, i. e. Radlex terms [1], to the affected areas within the images. Eventually, the relevant bits of information are then processed by the backend and made persistent for later reference.

We used a client-server-based approach, where each part that involves more computing is runnable on a separate platform. Note that regardless of the introduced modalities, all these approaches share the same underlying goal, namely to increase usability of the doctor's working routine environment, i. e., making knowledge acquisition fast and easy by providing state-of-art user interfaces to human medical expertise [7].

In this work we combine our novel interaction designs with current achievements concerning novel input devices, thus allowing for experiments with new interaction paradigms. For instance we have integrated an innovative mobile display systems in a form of a head-mounted display (Brother's AiRScouter WD-100G) which provide new ubiquitous possibilities for real-time interaction. We applied our technology to the HMD to provide mobile augmented reality interaction system for doctors that can be used during patient examination in the medical routine. The augmented reality system comprises of a digital smart pen (see Figure 1a) and a speech recognizer as input device. On the other side we have applied a see-through HMD (see Figure 1b) for visual feedback and a speech synthesis for audio feedback. See-through HMDs are of special interest as the doctors can focus on the patient during the examination process as the information is augmented in the view field.

In the following sections, we will describe the technical aspects with respect of the pen-based annotation framework, then we will also illuminate the relevant parts of our dialog system. Finally, we will highlight the interesting effects we gain when observing the interplay of both input channels combined with our new augmented reality approach.

MOBILE FEEDBACK PEN-BASED SYSTEM

Possible annotations on the printed paper may include terms to classify diseases which are formalised by using the International Classification of Diseases (ICD-10), annotations of regions-of-interest (ROI) in images, or pen gestures to choose predefined terms (e.g., anatomical concepts). In any case, the system maps the handwriting recognition (HWR) output to one or more medical concepts. Each of these ROIs can be annotated with anatomical concepts (e.g., *lymph node*), with information about the visual manifestation of the anatomical concept (e.g., *enlarged*), and/or with a disease category using ICD-10 classes (e.g., *Nodular lymphoma* or *lymphoblastic*). However, any combination of anatomical, visual, and disease annotations is allowed and multiple annotations of the same region are possible. Whenever an annotation is recognised and interpreted, the result is instantly augmented in the HMD (see Figure 1d) to give immediate feedback. Furthermore, the paper sheets contain special action areas to trigger additional functions of the system, such as a text-to-speech engine to provide optional audio feedback of the recognition results.

The architecture illustrated in Figure 2 is based on the pen component of the Touch&Write framework [1] and provides an overview of the proposed system. First, an ink collector component collects the online stroke data from the digital smart pen via Bluetooth. Now, the mode detection component [6] analyses the stroke information and classifies the annotation into different modes of the handwritten input.

In our scenario the system has to decide whether it deals with handwritten text information, image annotations, or pen gestures. The handwriting is analysed by using the MyScript engine of Vision Objects [2] and gesture recognition is performed by the iGesture framework [3]. Depending on the classification result, either the handwriting recognition or the pen gesture analysis is triggered. The recognition results are passed on to the interpretation layer via the event manager component. As the interpretation layer has a semantic mapping of the paper sheet layout, pen gestures and recognised texts are interpreted in the context of the diagnostic field where they occur.

Finally, a visualisation layer is responsible for providing an appropriate visualisation of the recognised diagnostic findings which depends on the hardware capabilities of the used HMD technology (see Figure 1d). The visualisation in the HMD mainly depends on the resolution of the HMD. In our demo prototype, Brother's AiRScouter (WD-100G) has been used with a screen resolution of 800x600 pixels. Due to the limited screen space, only selected parts of the complete finding form can be presented to the doctor (e.g., only the results of the MRT, see Figure 1d). Making a virtue out of necessity, we display only the diagnostic area of the actual form filling process, which does not overload the screen. For all areas, we present the schematic image with the marked ROIs combined with the selected anatomical concepts, as well as recognised handwritten annotations.

In summary, the mode detection of our smart pen automatically chooses the next step of analysis. The next analysis step will be, either:

- Pen gestures - for triggering predefined functionalities in our multi-modal system,
- Handwriting (common language combined with medical terms) - capturing diagnostic findings,
- Visual annotation - for marking ROIs in medical images.

Finally, as the smart pen "*knows*" the content of the medical forms and provide further information on the HMD, trigger actions, or provide a digitialized report of the finding.

[1] `http://www.radlex.org/` : Last seen 02/09/2013

[2] `http://www.visionobjects.com/en/myscript/about-myscript/` : Last seen 02/09/2013

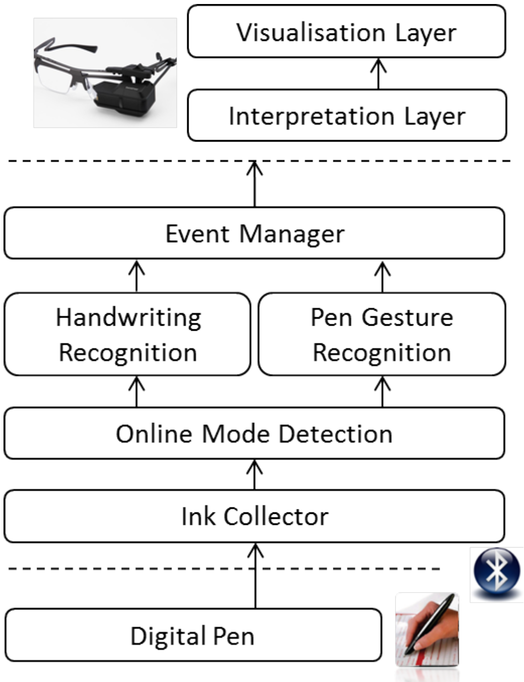
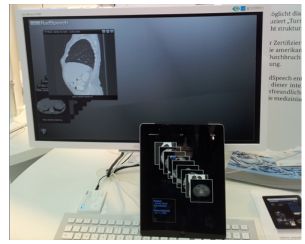
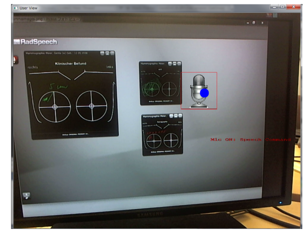


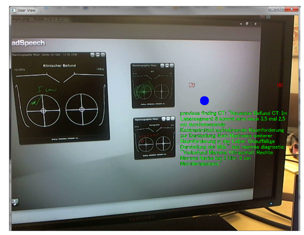
Figure 2: Proposed recognition and feedback data flow. The pen data is collected and further analysed in a three-tier architecture.



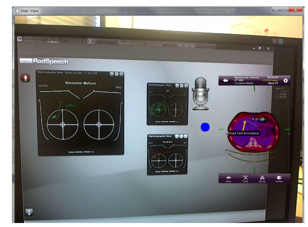
(a) The mobile medical diagnosis working station combined with a screen installation.



(b) The activation of the speech recognition is overlayed into the view of the user.



(c) Overlay of the patients file information.



(d) A video is streamed into the users sight.

Figure 3: The combination of a speech-based interface and a see-through interface.

THE SPEECH INTERFACE

We use the microphone array that is integrated into the Microsoft Kinect [3] to transmit audio signals to the speech recognition (Nuance Recognizer 9.0) [4]. The recognizer runs on a speech server that integrates also a speech synthesizer (Nuance SVOX). Our dialogue platform is based on ontological concepts that during runtime models the interaction process inside a production rule engine [2].

We have adopted a design direction that allows the activation of the dialog system using different type of devices. As a result, the user is able to choose the modality which is most convenient in a specific situation. Figure 3b shows the graphical output in the sight of the user when opening the microphone by the movement of the eye. The blue dot within the frame that contains the microphone icon represents the visual focus. The highlighted frame and a notification signalize the activation of the speech recognition to the user.

AN INTERACTION EXAMPLE

The following dialog demonstrates a real-world example; while the radiologist is analysing medical images on the screen application, he or she is requesting for additional information about the patient:

1 The doctor opens microphone using either eye gaze or pen gestures.
2 **Doctor says:** "Show me the previous finding in the HMD."
3 **HMD:** The sight of the doctor is augmented with the corresponding patient file.
4 **TTS:** "Previous Finding:. . . "
5 The doctor continues with the form-filling process.
6 **Doctor uses pen:** The Radlex terms *round*, *smooth*, *homogeneous* are marked.
7 **TTS:** "The annotation *round*, *smooth*, *homogeneous* has been recognized"

Figure 4 visualizes the interplay of the components given the dialog example in the context of the multi-device infrastructure. In the centre of the infrastructure we have a proxy, that is responsible to route and forward method invocation to any target recipient [5], i. e., I/O device. The eye tracker first interprets the gaze gesture as an open microphone command (see 1). The invocation of the actual method on the target is passed on through the network by means of the proxy. The calculation of the intended meaning of the speech input is done by the dialog system and will in turn result in the invocation of a remote call on the HMD part. Figure 3c shows the effect in the HMD of a `displayText` call that is triggered by the utterance (see 2). Simultaneously, the dialog system processes the audio format conveying the corresponding information (see 4). Further, during the form-filling process using the digital pen (see 5), user feedback is realized through multimodal output that involves the dialog system and the HMD. In particular, annotating ROI with text information is accompanied by audio feedback, note the sentence produced by the speech synthesizer (see 7) as the Touch&Write Framework recognizes the Radlex terms (see 6). The pen interface accesses the synthesis mechanism via the call `updateFindings` that is

[3] **http://www.microsoft.com/en-us/kinectforwindows/** : Last seen 02/09/2013
[4] **http://www.nuance.com/** : Last seen 02/09/2013
[5] **http://delight.opendfki.de/**: Last seen 02/09/2013

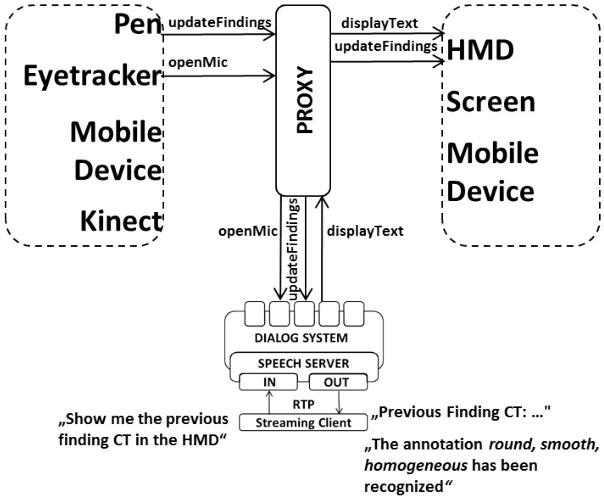


Figure 4: Any of the input devices on the left may serve to activate the speech modality, while on the right basically all output devices are available to give feedback to the user.

forwarded by the proxy to the dialog system. At the same time, the identical call effectuates real-time visual feedback augmenting the view of the doctor by displaying the selected terms in the see-through device (see the red coloured terms in Figure 1d). Besides the purpose of generating adequate multimodal feedback on the basis of the "digital pen print", we can also use the pen input to trigger explicit commands that goes beyond the context of the form-filling process. In Figure 3d a video is shown inside the augmented view. A `videoPlayback` call is triggered when underlining predefined terms within a designated area on the form. Based on our infrastructure we can easily make the latter functionality accessible also to other input modalities, such as speech. Finally we are able also to route the video stream to other output devices, such as the screen installation.

CONCLUSION

We presented a novel mobile real-time smart pen feedback environment which directly projects the results of the digital form-filling process into the eyes of the doctors. Radiologists can perform diagnostic reporting tasks by using their standardised form sheet and handwritten comments as well as simple pen annotations. To access additional information about the patient, we integrate a state of the art dialogue system called RadSpeech.

Our prototype employs modern HMD technology for displaying the real-time recognition results; as a result, our interaction system is mobile and the doctor does not need any additional devices, such as a tablet or smartphone, to check the digital version of the diagnosis. Moreover, to improve the robustness of the speech recognition in a real world scenario we used either pen or eye-gaze gestures to control the speech recognition instead of using continuous speech recognition.

Finally, after controlling the results, the digital version of the diagnostic finding can be transmitted to the radiology information system. This improves the quality and consistency of reports as well as the user interaction. Radiologists are also not forced to dictate information in the order in which it appears in the report. Most importantly, complete reports are available in seconds due to the mobile data acquisition and real-time feedback functionality.

ACKNOWLEDGMENTS

This work is part of THESEUS-RadSpeech (see **`www.dfki.de/RadSpeech/`**), supported by the German Federal Ministry of Economics and Technology (01MQ07016) and Nuance Foundation (see **`www.nuancefoundation.org/`**).

REFERENCES

1. M. Liwicki, M. Weber, T. Zimmermann, and A. Dengel. *Seamless Integration of Handwriting Recognition into Pen-Enabled Displays for Fast User Interaction*, pages 364 – 368. 2012.
2. N. Pfleger and J. Schehl. Development of advanced dialog systems with PATE. In *Proceedings of Interspeech 2006—ICSLP: 9th International Conference on Spoken Language Processing, Pittsburgh, PA, USA*, pages 1778–1781, 2006.
3. B. Signer, U. Kurmann, and M. Norrie. iGesture: A General Gesture Recognition Framework. In *9th International Conference on Document Analysis and Recognition 2007.*, volume 2, pages 954 –958, sept. 2007.
4. D. Sonntag, M. Liwicki, and M. Weber. Digital pen in mammography patient forms. In *Proceedings of the 13th international conference on multimodal interfaces (ICMI '11).*, pages 303–306, Alicante, Spain, 2011. ACM, New York, NY, USA.
5. D. Sonntag and C. H. Schulz. Multimodal multi-device discourse and dialogue infrastructure for collaborative decision making in medicine. In *Proceedings of the International Workshop on Spoken Dialogue Systems Technology*, IWSDS12. Springer, 2012.
6. M. Weber, M. Liwicki, Y. T. H. Schelske, C. Schoelzel, F. Strauss, and A. Dengel. MCS for Online Mode Detection: Evaluation on Pen-Enabled Multi-Touch Interfaces. In *11th International Conference on Document Analysis and Recognition*, 2011.
7. D. L. Weiss and C. Langlotz. Structured reporting: Patient care enhancement or productivity nightmare? *Radiology*, 249(3):739–747, 2008.