RadSpeech's Mobile Dialogue System for Radiologists

Daniel Sonntag, Christian Schulz, Christian Reuschling and Luis Galarraga
German Research Center for AI - DFKI
Stuhlsatzenhausweg 3
Saarbrücken, 66123, Germany
Firstname.lastname@dfki.de

ABSTRACT

With RadSpeech, we aim to build the next generation of intelligent, scalable, and user-friendly semantic search interfaces for the medical imaging domain, based on semantic technologies. Ontology-based knowledge representation is used not only for the image contents, but also for the complex natural language understanding and dialogue management process. This demo shows a speech-based annotation system for radiology images and focuses on a new and effective way to annotate medical image regions with a specific medical, structured, diagnosis while using speech and pointing gestures on the go.

Author Keywords: Speech Dialogue, Mobility, Healthcare

ACM Classification: H.5.2 [Information Interfaces And Presentation]: User Interfaces - Interaction styles;

General terms: Management, Reliability, Human Factors

INTRODUCTION

In radiology, computed tomography (CT) or magnetic resonance imaging (MRI) images are used to both diagnose and treat diseases visualised within the human body. Clinical research partners, such as the University Hospital Erlangen in Germany, produce a lot of patient images and have a total of about 65 TB of medical images. The problem with current technology is that a clinician cannot directly create a structured report while scanning the images: in this *eyes-busy* setting, he or she can only dictate the finding to a tape-recorder. After the reading process, he or she can replay the dictation to manually fill out a patient's finding form or delegate other personnel. But since the radiologist has to check the form again, task delegation does not save time. With our technology developed over the last 5 years, we implemented the first mobile dialogue system on the iPad and iPhone, which is tuned for the radiology domain and makes the complete system unique. Our solution not only provides more robustness compared to speech-to-text systems (we use a rather small, dedicated, and context-based speech grammar which is also very robust to background noise), it also fits very well into new radiology reporting processes which will be established in Germany and the U.S. over the next several years: in *structured reporting* you directly have to create database entries instead of text.

RadSpeech (www.dfki.de/RadSpeech/) is the design and implementation of a multimodal dialogue system for structured radiology reports. With traditional user interfaces, users may browse or explore patient data, but little to no help is given when it comes to the interpretation of what is being displayed or structuring the input. In order to allow a radiologist to annotate special image regions in a database format and to search for similar cases, we developed a Desktop-based manual annotation tool called RadSem [3]. Anatomical structures and diseases can be annotated while using the auto-completion combo-boxes with a search-as-you-type functionality. The resulting annotation is accurate but very time-consuming.

In addition, RadSem did not fulfill the special requirement that clinicians have: to have access to a coherent view of image data within their particular diagnosis or treatment context in the radiology department while they are skimming many image series and thousands of pictures in a minute's time. Although it is widely reductive to put it this way, a senior radiologist has three main goals: (1) access the images and image (region) annotations, (2) complete them, and (3) refine existing annotations. We argued that these tasks can best be fulfilled while using a multimodal dialogue system, and first we experimented with a large touchscreen installation [1, 2]. Since the results with speech interaction were very promising, we tried to opt for the mobile context. Our mobile system is expected to provide the radiologist with the ability to review images when outside the laboratory, and to make diagnosis without having to be back at the workstation,
see www.youtube.com/watch?v=uBiN119_wvg.

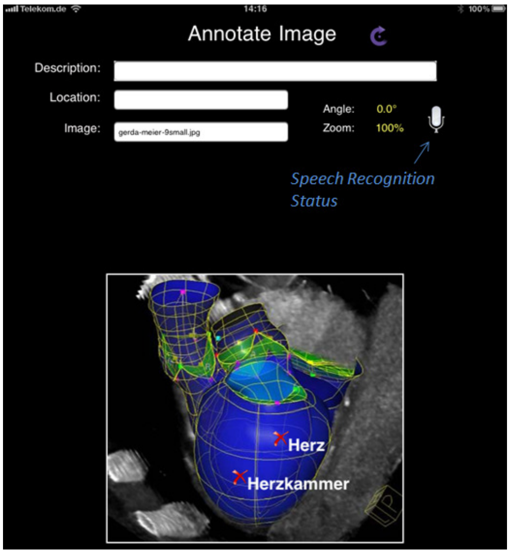
The semantic dialogue system presented in this demo should be used to ask questions about the image annotations while engaging the clinician in a natural speech dialogue. Different semantic views of the same medical images (such as structural, functional, and disease aspects) can be explicitly stated, integrated, and asked for. This is the essential part of the knowledge acquisition process [3]. Two aspects are implemented and shown in the demo. First, the inspection of and navigation through the patient's data, and second, the annotation of radiology images by use of speech and gestures.

PREPRINT
IUI'12, February 14–17, 2012, Lisbon, Portugal.
ACM 978-1-4503-1048-2/12/02.

EXAMPLE DIALOGUE

A radiologist treats a lymphoma patient. The patient visits the doctor after chemotherapy for a follow-up CT examination. The focus of the speech-based interactions is the following sub-dialogue (for simplicity, the cancer annotation is replaced by a simple anatomy annotation).

- U: “Show me the CTs, last examination, patient XY.”
- S: Shows corresponding patient CT studies as DICOM picture series and MRI images and MRI videos.
- U: “Annotate this picture with ‘Heart’ (+ pointing gesture) and ‘Heart chamber’” (+ pointing gesture)
- S: Shows the new annotations on the image and confirms a database update.



Telekom.de 14:16 100%

Annotate Image

Description:

Location:

Image: gerda-meier-9small.jpg

Angle: 0.0°

Zoom: 100%

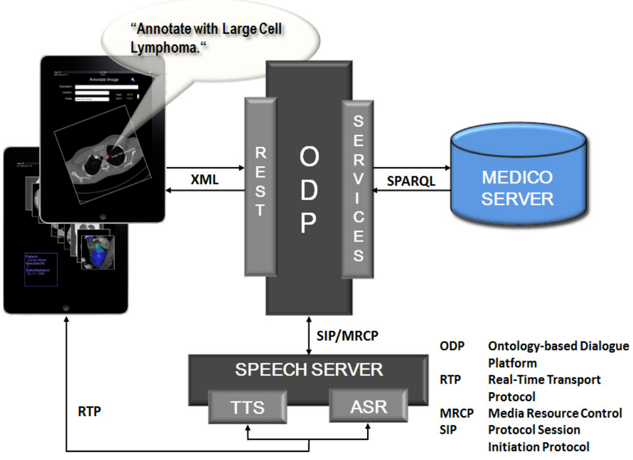
Speech Recognition Status

Figure 1. Screenshot of the Speech-based Image Annotation Screen

Figure 1 shows the screenshot of the annotation screen. Upon touching a region in the white square, the speech recognition system is activated. After recognition, the speech and gesture modalities are fused into a complex annotation using a combination of medical ontologies. For disease annotations for example, the complete Radlex (http://www.radlex.org/) terminology can be used. More complex interactions can be seen in the demo video at http://www.youtube.com/watch?v=uBiN119_wvg, where a set of multi-touch gestures to control the image selection and manipulation phase without the usage of distracting screen buttons is shown as well.

SYSTEM ARCHITECTURE

We implemented a middleware that invokes the MEDICO backend server directly. This middleware is implemented by means of the Ontology-based Dialogue Platform (ODP): the central *Event Bus* is responsible for routing message between the *Dialogue System* and other connected components. Such components include a speech recogniser (*ASR*) and a speech synthesis (*TTS*) module. On the client, e.g., an iPad, only a slim application is needed to encode the speech and gesture input and receive the speech synthesis to be played (Figure 2, details are explained in [2]). New clients can be added easily by, e.g., downloading an iPad app.



“Annotate with Large Cell Lymphoma.“
XML
REST
ODP
SERVICES
SPARQL
MEDICO SERVER
SIP/MRCP
SPEECH SERVER
TTS
ASR
RTP
ODP Ontology-based Dialogue Platform
RTP Real-Time Transport Protocol
MRCP Media Resource Control
SIP Protocol Session Initiation Protocol

Figure 2. Technical Architecture of RadSpeech

CONCLUSION

Today, medical images have become indispensable for detecting and differentiating pathologies, planning interventions, and monitoring treatments. We have presented a mobile semantic speech dialogue system for the radiologist. Our new prototypical dialogue system provides the radiologist with the ability to review images when outside the laboratory, to annotate important image regions while using speech and gestures, and to make a *structured* diagnosis without having to be back at the workstation. Our experts predict that mobile dialogue technology will dominate mobile healthcare in just a few short years. The speech-based dialogue system RadSpeech is currently a part of a larger clinical study about the acquisition of medical image semantics at Siemens Healthcare, the University Hospital in Erlangen, and the Imaging Science Institute (ISI).

ACKNOWLEDGEMENTS This research has been supported by the research programme THESEUS, which is funded by the German Federal Ministry of Economics and Technology under the grant number 01MQ07016.

REFERENCES

1. Cohen, P.R., Oviatt, S. The role of voice input for human-machine communication. Proc. Natl. Sci. USA (1995)
2. Sonntag D., Reithinger N., Herzog G. and Becker T. A Discourse and Dialogue Infrastructure for Industrial Dissemination. Proceedings of IWSDS, Lee et al. (Eds.): LNAI 6392, Springer (2010)
3. Sonntag, D. Intelligent Interaction and Incremental Knowledge Acquisition for Radiology Images. Proceedings of SAMT, Springer (2010)