A Speech-based Dialogue Platform for Supporting Medical Image Analysis

Christian Husodo Schulz and Daniel Sonntag

DFKI GmbH,
Stuhsatzenhausweg 3, 66123 Saarbrücken, Germany
{christian.husodo.schulz,daniel.sonntag}@dfki.de
http://www.dfki.de/RadSpeech/

Abstract. The idea behind the RadSpeech platform is to provide a repository of computer-aided tools supporting diagnostic image analysis with the goal to ease and support the work of the medical expert on a daily basis. We focus on state-of-the-art interaction paradigms in a radiology related working environment. In particular, we have implemented a semantic speech dialogue system for radiologists which has been deployed in diverse scenarios. With traditional user interfaces little support is provided when it comes to the interpretation of visualized patient data. The speech based dialogue platform brings together meta-information of patient data (i. e., patient images) and the ontological representation of knowledge used by the dialogue system. As a result, patient data is accessible to the clinician while using natural language as the primary communication mode.

1 Introduction

A word about clinical care: there is hardly any other application area where there is such an amount of data that need to be examined, analyzed, and organized by humans in a daily routine. Information about a patient comprises health records, laboratory records, and medical images. For the medical expert it takes great effort to cope with the flood of data and structure them to medical reports [1]. We claim that there is a major demand for a more efficient handling with medical (image) data, especially when it comes to annotating specific body regions with medical concepts. Recent works on this issue suggest to enhance medical data with semantic technology in order to make the resulting image annotations searchable and comparable [2]. Semantic knowledge on top of medical data even enables computer-aided reasoning processes and allows for interconnections among patient data beyond ordinary links from patient to raw data. This paper describes our speech-based platform for supporting medical analysis, namely the knowledge acquisition process. More precisely, it enables the medical expert to attach meta-information to medical images without disrupting the workflow.

In what follows, we introduce the underlying fundament of our multimodal dialogue system, the RadSpeech Platform. In the subsequent section we provide a detailed description of the semantic annotation methodology, which is the process of combining concepts from a medical ontology and medical patient image data through natural interaction (the human voice) in combination with touch gestures on, e. g., mobile interaction devices.

2 RadSpeech Platform

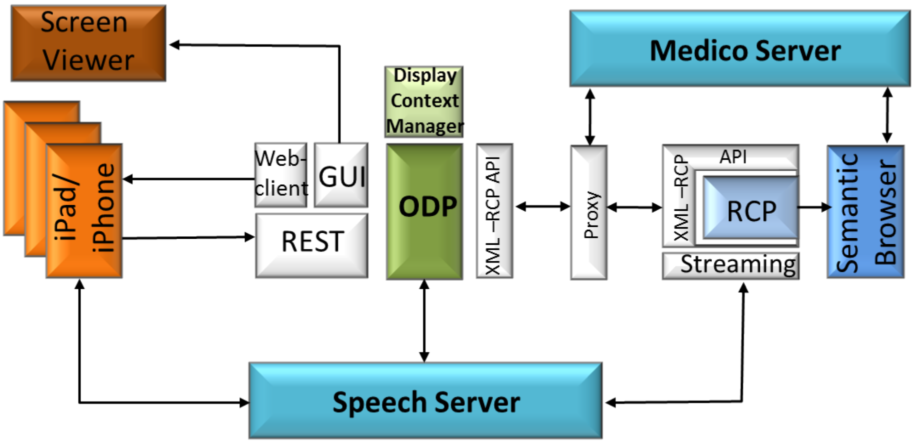


Fig. 1. Complete ecosystem of the speech-based platform for medical findings

Figure 1 shows the heart of the dialogue system platform. We use a modified version of the ODP, *Ontology-based Dialogue Platform* [3, 4], which is based on established ontology standards [5, 6]. One of the core modules of the dialogue system is a so-called rule engine PATE, *a Production rule system Based on Typed Feature Structures* [7, 8]. In addition, we attached several interface components to the ODP which enable the communication to the different types of peripheral devices. We will present the Display Context Manager in more detail in the section 3; it maintains and manipulates the medical data during runtime. This is the component where semantic annotations (performed via voice) are bound to the medical images and image regions. Another important component is the speech server that is integrated into the platform, wherein the Nuance Recognizer 9.0 for speech recognition and SVOX for speech synthesis can be used in a distributed client-server architecture. In terms of responsiveness and robustness these software products have already proven their suitability in a variety of most recent implementations[9–11]. Especially in the mobile setting, where it is impossible to control the conditions of the background noise, it is indispensable to have reliable speech processing components (as well as dedicated speech grammars) as one of the major requirements for implementing industry-relevant spoken dialogue system processing chains.

Based on the ODP platform, we have implemented several input and output modules of a discourse and dialogue infrastructure for industrial dissemination, such as those the medical domain [4]. The ecosystem outlines all the components and modules that have been implemented and realized so far in the context of the RadSpeech dialogue system for radiologists. The platform has been designed in order to ease its adaption according to new applications. The applications which run on (mobile) clients are grouped into three different ”work assignments” in the context of the tool support for medical image analysis and respective patient finding processes:

1. Mobile applications, i. e., on the iPad, iPhone
2. Viewer for collaborative findings, i. e., RadComet (a big interaction screen, explained below)
3. RCP-Based Application (Rich Client Platform[1]) for the desktop environment, i. e., Planar Viewer
4. Web-based Semantic Search on patient data, i. e., Semantic Browser

The RadSpeech platform enables the user to interact with the system via voice and deictic input, in single mode but also in mixed multimodal mode with the peripheral devices, e. g., the iPhone or iPad. Each client that is included in the ecosystem is assigned to a unique session and is managed within the platform in a separate runtime environment. Thus, medical image manipulation of one user does not affect the analysis of other users that are connected to the dialog server. At a later stage when the user decides to store the changes, commitments on data manipulation can be sent to the Medico Server. The main purpose of the Medico Server is the persistent patient data storage. During the editing stage, the patient data is encapsulated inside a user session and is associated to one peripheral device. However, the platform also permits the exchange of patient data among the sessions. Dedicated user commands may propagate information about the current patient to other medical experts. A touchscreen surface plane can display the outcome of the analysis following specific logical arrangements of the design elements [12]. Combining the big screen—the so-called RadComet—with the mobile devices, we have created a scenario where relevant parts of the medical findings can be shared on a big screen (in the case that the expert is willing to share his patient-related data). In the following, we will set the focus on the main interaction sequence, namely the process of annotating image by speech commands; this interaction sequence (for the purpose of knowledge acquisition) is covered by all prototype applications of the RadSpeech platform. We expect a major benefit in terms of utility (speed and comfort) using speech technology to label image regions with medical terms. The medical labels denote RadLex terms, of which there are more than six thousands covered by the official RadLex repository [13]. Our approach tries to meet the new radiology reporting process that foresees the use of a vocabulary based on a defined taxonomy, e. g., RadLex.

[1] http://wiki.eclipse.org/index.php/Rich_Client_Platform

3 The Semantic Medical Image Annotator

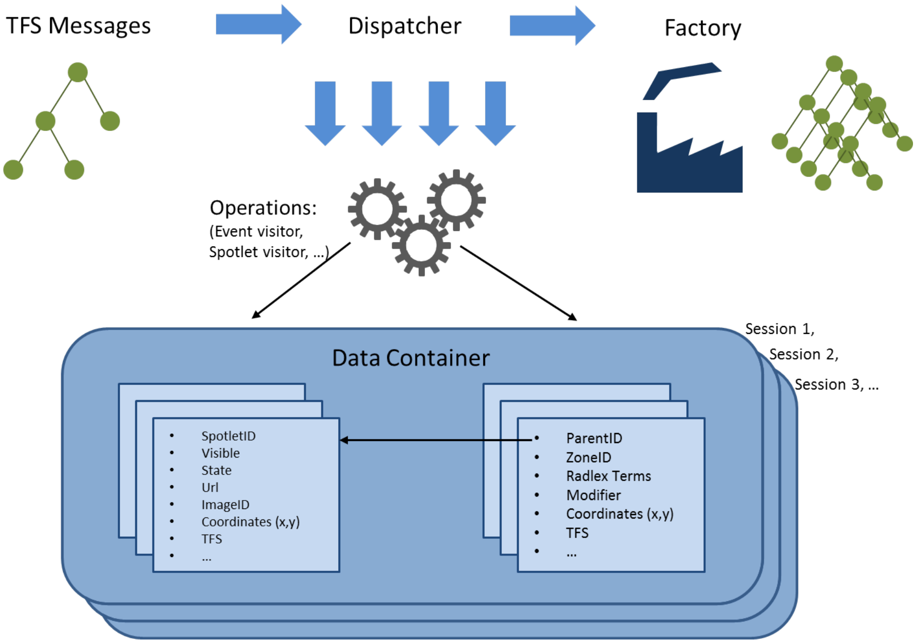
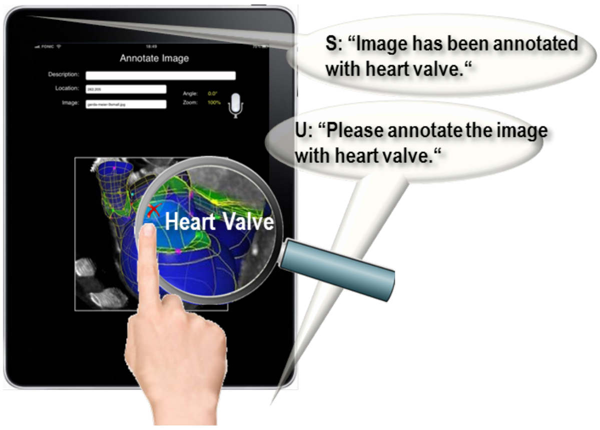


Fig. 2. Display Context Manager which separates operational behavior from its data

In the course of extending the RadSpeech platform with different clients that run in parallel, we have designed and implemented a Display Context Manager that is running inside the dialogue system and is responsible for maintaining the internal (data) state and for invoking behavioral operations on the patient data, medical images in particular. The retrieved patient data that stem from the Medico Server are translated into ontological instances and are available to be manipulated in the working memory during a session. The Display Context Manager is in charge of dispatching the command messages which are also ontological instances, i. e., *Typed Feature Structures* (TFS) [14]. In the next step, the corresponding TFS is handed over to proper operational components possessing exclusive access rights to write on medical data as shown in figure 2. The working data is located inside a container, maintaining so-called *spotlets* and *zones*. Spotlets are also container for patient images pointing to various useful meta-information (e. g., Dicom meta data about the image recording process in the hospital such as date, time, image modality, and the patient's name). Zones are containers administrating the annotations associated with the spotlets. After the operation has been executed on the data in terms of updating their state, the dispatching mechanism invokes an adequate factory method that generates an ontological instance wrapping the new and modified state of spotlets and zones for further processing. In the following, we will learn how the manipulation of a medical image from the ontological view, based on the speech utterance "*Please annotate the image with heart valve*," is performed.

3.1 Manipulating meta-information on images

The user turn of manipulating annotations on medical images is preceded by several interactions combined by voice and deictic input. First, we will introduce the context of the dialogue scenario before explaining the details on the ontological level. At the beginning of the interaction the user has to retrieve the data of a patient (by voice or gesture), which is *Gerda Meier*[2] in this example. Then, the user has to detect the medical image he wants to edit, which is the heart, see the related image id *image_1_GerdaMeier* in the figure 5 in our example. After confirming the selection by a double tap on the respective image thumbnail, the state as presented in the screenshot in figure 3 has been accomplished. Essentially, the user is able to activate the microphone by a simple tap in order to provide the annotation by a speech utterance.



Annotate Image

Description:

Location: 262,205

Image: gerda-meier-5small.jpg

Angle: 0.0°

Zoom: 100%

S: "Image has been annotated with heart valve."

U: "Please annotate the image with heart valve."

Fig. 3. Visualization of the user turn that represents the medical image annotation on the iPad

The TFS message (figure 4, on the right) invokes the Display Context Manager to change the annotation state of a spotlet. The Display Context Manager is able to identify the correct spotlet and zone by interpreting former deictic input. Each relevant gesture event on the iPad is translated into a *radspeech#ImageInputEvent* message (see figure 4, on the left), which is further processed by the Display Context Manager, i. e., the *Event Visitor*. Concerning

[2] The patient names published here are pseudonyms.

```
1  <object type="radspeech#ImageInputEvent">
2    <slot name="odp#hasContent">
3      <object type="medico#ImageAnnotation">
4        <slot name="odp#isSelected"/>
5        <slot name="medico#annotation"/>
6      </object>
7    </slot>
8    <slot name="odp#action">
9      <value type="String">
10         <![CDATA[select_zone]]>
11     </value>
12   </slot>
13   <slot name="radspeech#id">
14     <value type="String">
15         <![CDATA[1]]>
16     </value>
17   </slot>
18   <slot name="comet#xCoordinate">
19     <value type="Float">
20         <![CDATA[252]]>
21     </value>
22   </slot>
23   <slot name="comet#yCoordinate">
24     <value type="Float">
25         <![CDATA[190]]>
26     </value>
27   </slot>
28 </object>
```

```
1  <object type="medico#AnnotateTask">
2    <slot name="odp#hasContent">
3      <object type="medico#MedicoSpotlet"/>
4    </slot>
5    <slot name="odp#hasContent">
6      <object type="medico#ImageAnnotation">
7        <slot name="medico#annotation">
8          <value type="String">
9            <![CDATA[herzklappe]]>
10         </value>
11       </slot>
12     </object>
13   </slot>
14   <slot name="medico#linked">
15     <object type="medico#Modifier">
16       <slot name="radspeech#modifier">
17         <value type="String">
18           <![CDATA[add_ann]]>
19         </value>
20       </slot>
21     </object>
22   </slot>
23 </object>
```

Fig. 4. The annotation process represented by ontological instances

image annotations, the Display Context Manager is informed of the coordinates for a potential new zone container by the Event Visitor. Given the following *medico#AnnotateTask* message, first an operation located on a level with access to manipulate zone containers, then a so-called *Zone Visitor* is triggered. In case that a zone with same coordinates has not been created, a fresh zone with the newly registered coordinates is instantiated, thereby binding the RadLex term “Herzklappe” (see line 9 on the right of figure 4). In a second variation, the term will be added to an already existing annotation. Eventually, the created zone points to the spotlet where it is located. This example shows how multiple modalities, here a pointing gesture together with a speech-based input, are successfully fused into one XML-based representation.

3.2 Meta-information of medical images

Besides the exchange of ontological instances within the rule engine, we have defined an own XML-based presentation markup language (PREML) which preserves meta-information that stems from the display context (data related to the analysis of the interaction are supposed to be handled by the client). The PREML leaves the tasks that are concerned with the appearance of the graphical elements underspecified and provides the flexibility for client-based applications to define their own (visual) characteristics on the graphical surface layer.

As for medical image annotations, the most relevant client-independent meta-information of medical images includes (1) the RadLex term itself, (2) the so-called annotation-*modifier*, (3) the coordinates of the *zone* to be annotated, and (4) a unique identifier assigned to the zone. The PREML structure in figure 5 encodes this meta-information and adds the distinctive input given by the

```
<preml−radspeech:presentation
    xmlns:preml−radspeech=”http://www.dfki.de/markup/preml−radspeech”
    preml−radspeech:id=”image_1_GerdaMeier”
    preml−radspeech:target=”radspeech”
    preml−radspeech:type=”update”
    preml−radspeech:x=”1138.0”
    preml−radspeech:y=”208.0”>

...

  <preml−radspeech:zones>
    <preml−radspeech:zone
        preml−radspeech:annotationLabel=”Herzklappe”
        preml−radspeech:height=”10.0”
        preml−radspeech:id=”zone_2”
        preml−radspeech:modifier=”add_ann”
        preml−radspeech:new=”true”
        preml−radspeech:width=”10.0”
        preml−radspeech:x=”253.0”
        preml−radspeech:y=”253.0” />
  </preml−radspeech:zones>
</preml−radspeech:presentation>
```

Fig. 5. Presentation messages (PREML) containing meta-information about the annotation of a patient image

medical expert after the process of annotating the images (processed by the Display Context Manager).

4 Conclusion

We have presented a platform for radiologists that aims to ease the process of annotating medical images with semantic concepts along the daily routine. Natural speech synthesis in combination with robust automatic speech recognition are our success factors, embedded into a pleasant HCI experience by using a mobile interaction device (also see `http://www.youtube.com/watch?v=uBiN119\_wvg`). We hypothesize that the impact of semantically enhanced medical data will become significant in the future because then every single medically-annotated image becomes a part of a knowledge base where reasoning processes can operate on. Several research teams have just begun to assemble specific reasoning strategies in the medical (imaging) domain [15].

The benefit of interlinked medical data becomes apparent as soon as not only patient data of one patient and one expert is considered, but the data of several similar patients and an expert collective. Our current effort concentrates on the elaboration of collaborative scenarios where a number of experts will be able to take part in the process of the analysis and annotation of shared (image) data. We expect an extra boost for the purpose of digitizing medical data and annotating concepts of a medical ontology; therefore we are currently investigating in implementations of more detailed collaborative scenarios that meet the requirements of an expert collective.

References

1. D. L. Weiss and C.P. Langlotz. Structured reporting: Patient care enhancement or productivity nightmare? *Radiology*, 249(3):739–747, 2008.
2. Daniel Sonntag, Pinar Wennerberg, and Sonja Zillner. Applications of an ontology engineering methodology accessing linked data for medical image retrieval. In *Proceedings of the AAAI Spring Symposium "Linked Data meets Artificial Intelligence". AAAI Spring Symposium, March 22-24, Stanford,, CA, United States.* Stanford University, 2010.
3. Norbert Pfleger. FADE - An Integrated Approach to Multimodal Fusion and Discourse Processing. In *Proceedings of the Dotoral Spotlight at ICMI 2005*, Trento, Italy, 2005.
4. Daniel Sonntag, Norbert Reithinger, Gerd Herzog, and Tilman Becker. A discourse and dialogue infrastructure for industrial dissemination. In Gary Geunbae Lee, Joseph Mariani, Wolfgang Minker, and Satoshi Nakamura, editors, *IWSDS*, volume 6392 of *Lecture Notes in Computer Science*, pages 132–143. Springer, 2010.
5. Frank Manola and Eric Miller. RDF primer. W3C recommendation, W3C, February 2004. Published online on February 10th, 2004 at `http://www.w3.org/TR/2004/REC-rdf-primer-20040210/`.
6. Deborah L. Mcguinness and Frank van Harmelen. Owl web ontology language overview, February 2004.
7. Norbert Pfleger and Jan 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.
8. Daniel Sonntag and Christian Husodo Schulz. Monitoring and explaining reasoning processes in a dialogue system's input interpretation step. In *Proceedings of the Workshop on Explanation-Aware Computing (EXACT)*, 2011.
9. Christian Husodo Schulz, Ingo Zinnikus, Patrick Kapahnke, Jochen Frey, Robert Neßelrath, and Jan Alexandersson. Universally accessible interactive services on tv. In Reiner Wichert; Birgid Eberhardt, editor, *Ambient Assisted Living, 4. AAL-Kongress 2011.* VDE, Springer, 1 2011.
10. Jan Schehl, Alexander Pfalzgraf, Norbert Pfleger, and Jochen Steigner. The BabbleTunes System. Talk to Your IPod! In *Proceedings of the 10th International Conference on Multimodal Interfaces (ICMI)*, 2008.
11. Robert Neßelrath, Christian Husodo Schulz, Jan Schehl, Alexander Pfalzgraf, Norbert Pfleger, Verena Stein, and Jan Alexandersson. Homogeneous multimodal access to the digital home for people with cognitive disabilities. In *Ambient Assisted Living 2009. 2. Deutscher AAL-Kongress mit Ausstellung / Technologien - Anwendungen (AAL-09), January 27-28, Berlin, Germany.* VDE, 2009.
12. Daniel Sonntag and Manuel Möller. Prototyping semantic dialogue systems for radiologists. In *Proceedings of the Sixth International Conference on Intelligent Environments (IE)*, 2010.
13. Curtis P. Langlotz. Radlex: A new method for indexing online educational materials. *RadioGraphics*, 26:1595–1597, 2006.
14. Bob Carpenter. *The Logic of Typed Feature Structures.* Number 32 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, UK, 1992.
15. Sonja Zillner and Daniel Sonntag. Image metadata reasoning for improved clinical decision support. volume 22, pages 37–46. Springer, 2012.