New Business to Business Interaction: Shake your iPhone and speak to it.

Daniel Porta
German Research Center for AI
Stuhlsatzenhausweg 3
66123 Saarbruecken
Germany
daniel.porta@dfki.de

Daniel Sonntag
German Research Center for AI
Stuhlsatzenhausweg 3
66123 Saarbruecken
Germany
sonntag@dfki.de

Robert Neßelrath
German Research Center for AI
Stuhlsatzenhausweg 3
66123 Saarbruecken
Germany
Robert.Nesselrath@dfki.de

ABSTRACT

We present a new multimodal interaction sequence for a mobile multimodal Business-to-Business interaction system. A mobile client application on the iPhone supports users in accessing an online service marketplace and allows business experts to intuitively search and browse for services using natural language speech and gestures while on the go. For this purpose, we utilize an ontology-based multimodal dialogue platform as well as an integrated trainable gesture recognizer.

Categories and Subject Descriptors

H.5.2 [**HCI**]: User Interfaces – Dialogue, Voice I/O, Interaction styles

General Terms

Design, Experimentation, Human Factors

Keywords

Multimodal Interaction, Mobile Business Services, Usability, User Experience, Productivity

1. INTRODUCTION

Time is money. Today's field staff is normally equipped with mobile devices allowing them to ubiquitously access and interact with urgent business processes while on the go, without having to access their standard desktops and common interfaces. However, squeezing the frontend of a complex business application into a hardware-restricted mobile device inevitably raises questions that are hard to answer regarding the interaction design. Implementing a reliable, trustful, and usable mobile interface for decision support in business-critical situations is challenging. A functional mobile solution, similar to [3], is required to optimize the business processes. The approach we explore uses a multimodal mobile client for the iPhone, intended to ease the access to an emerging online marketplace for electronic business web services while on the go. In this contribution, we discuss a task-oriented multimodal interaction sequence that can be performed by the user on the mobile device in order to come to a decision and contribute to the business processes. We briefly discuss the dialogue-based interaction sequences and present the architecture of the overall system.

MobileHCI'09, September 15 - 18, 2009, Bonn, Germany.
ACM 978-1-60558-281-8.

2. INTERACTION SEQUENCE

The business scenario is as follows: the mobile client application, which is integrated into a company-internal Enterprise Resource Planning (ERP) system, allows a business expert of the company's purchasing department to constantly stay in touch with the most recent developments. This way, a delay in business processes can be avoided. In our scenario, the business expert has to check a purchase request for a service (offered in the service marketplace). He is interested in, e.g., quality and cost standards before the transaction can be carried out.

Figure 1 shows screenshots for a typical interaction sequence between the user and the system. Starting with a new ticket that needs the business expert's attention, the user opens it by tapping on the respective list entry. The resulting detailed description of the ticket is depicted in Figure 1 (a).

(1) **U:** *Moves the device closer to his mouth.*

(2) **S:** Automatically activates the microphone and waits for speech input. The microphone button turns green.

Turn **(1)** and **(2)** are performed before every speech interaction.

(3) **U:** *"Show me the price models of this [+ pointing gesture on the service name] service, please."*

(4) **S:** The service description is shown with a focus on the price model information. *"There are 2 price models."* (Figure 1 (b))

(5) **U:** Reads the description and *shakes* the device.

(6) **S:** Discards the service description and restores the former display context. (Figure 1 (a))

(7) **U:** *"Show me alternative services."*

(8) **S:** The retrieved services are shown in a list. *"I found 4 alternatives."* (Figure 1 (c))

(9) **U:** *"Sort the entries according to reliability."*

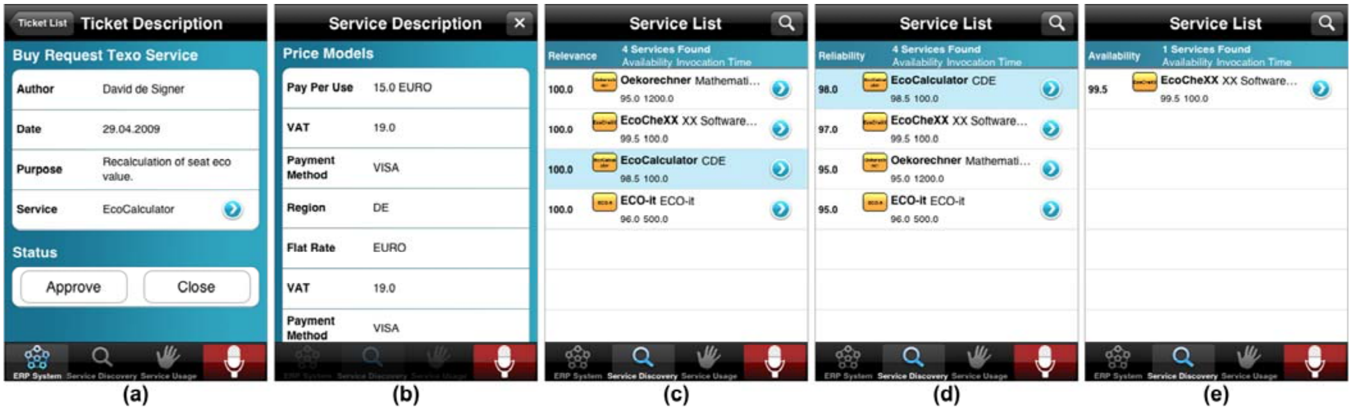
(10) **S:** *"I sorted the entries in descending order according to reliability."* (Figure 1 (d))

(11) **U:** *"Show only services with an availability rate better than 99%."*

(12) **S:** *"Only 1 service remains."* (Figure 1 (e))

(13) **U:** Does not agree to the remaining choice and *shakes* the device again.

(14) **S:** Removes the applied filter and shows the whole list again. (Figure 1 (d))



(a) Ticket Description

Buy Request Texo Service

Author David de Signer
Date 29.04.2009
Purpose Recalculation of seat eco value.
Service EcoCalculator

Status

(b) Service Description

Price Models

Pay Per Use 15.0 EURO
VAT 19.0
Payment Method VISA
Region DE
Flat Rate EURO
VAT 19.0
Payment Method VISA

(c) Service List

Relevance 4 Services Found
Availability Invocation Time

100.0 Oekorechner Mathemati... 95.0 1200.0
100.0 EcoCheXX XX Software... 99.5 100.0
100.0 EcoCalculator CDE 98.5 100.0
100.0 ECO-it ECO-it 96.0 500.0

(d) Service List

Reliability 4 Services Found
Availability Invocation Time

98.0 EcoCalculator CDE 98.5 100.0
97.0 EcoCheXX XX Software... 99.5 100.0
95.0 Oekorechner Mathemati... 95.0 1200.0
95.0 ECO-it ECO-it 96.0 500.0

(e) Service List

Availability 1 Services Found
Availability Invocation Time

99.5 EcoCheXX XX Software... 99.5 100.0

Figure 1: Screenshots of the Sample Interaction Sequence

In **(1)**, the user intuitively moves the device closer to his mouth since he intends to interact with the device using speech. He slightly tips the device since the microphone is placed at the lower edge of the device. This gesture is recognized by interpreting the sensor data of the built-in accelerometer. As a result, the dialogue system automatically opens the microphone **(2)**. The advantage over, e.g., other applications for the iPhone is that the user does not need to move his device to his ear to activate the microphone, allowing him to always see and refer to the contents displayed on the GUI, which is in fact required for real multimodal interaction. A disadvantage might be that interpreting accelerometer sensor data requires further computation (which is outsourced in our case, see chapter 3). Request **(3)** is answered by fusing the multiple input modalities; **(7)** is answered by inferring the missing pieces of information (to which service shall alternatives be found) from the discourse context. By shaking the device in **(5)** and **(13)**, the user triggers an “undo” command. For this purpose, an interaction history stack is maintained by the system. So, in general, all interactions, regardless of the input modality, can be undone. For example, if the last interaction activated the microphone (either by pushing the button or by performing the above described gesture) the user can deactivate it again by shaking the device.

If the user is finally convinced, he can return to the ticket description (Figure 1 (a)) and carry out the transaction.

3. ARCHITECTURE

The overall system architecture is depicted in Figure 2.

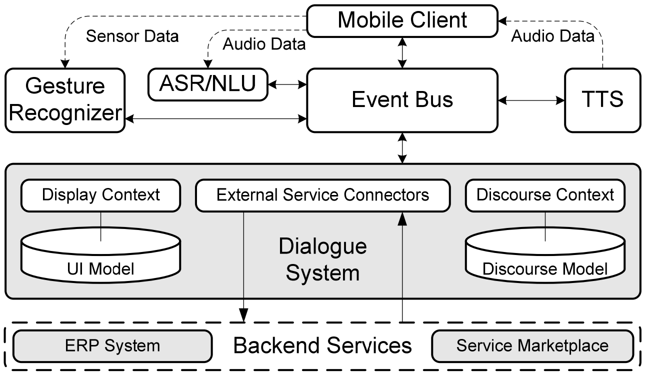


Figure 2: Overall System Architecture

We propose a dedicated middleware that invokes the backend services directly (in contrast to, e.g., [4], where services are adapted by providing simplified versions). This middleware is implemented by means of the Ontology-based Dialogue Platform (ODP) [2]: the central *Event Bus* is responsible for routing message between the *Dialogue System* and other connected components. Such components include a speech recognizer (*ASR*) and a speech synthesis (*TTS*) module. A trainable *Gesture Recognizer* [1] receives accelerometer sensor data from the *Mobile Client* and sends the interpretation to the *Dialogue System*, which observes and maintains the current *Display* and *Discourse Context*. It also provides interfaces (*External Service Connectors*) to relevant *Backend Services*. Speech and pen-based gestures are active user input modes. These can be used to issue commands in our B2B environment. While incorporating the iPhone’s accelerometer data, we included more passive input modes towards a perceptual mobile interface. Our next steps include more fine-grained multimodal fusion rules for the passive input modes, i.e., context-sensitive interpretations of the “shake”.

4. ACKNOWLEDGMENTS

This research has been supported by the THESEUS Research Programme in the Core Technology Cluster WP4 and the TEXO use case, which was funded by means of the German Federal Ministry of Economy and Technology (01MQ07012). The authors take the responsibility for the contents.

5. REFERENCES

[1] Neßelrath, R. and Alexanderson, J.: A 3D Gesture Recognition System for Multimodal Dialog Systems. In Proceedings of the 6th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue, Pasadena, USA, (2009)

[2] Schehl, J., Pfalzgraf, A., Pfleger, N., and Steigner, J.: The babbleTunes system: talk to your ipod!. In Proceedings of IMCI '08. ACM, New York, NY, (2008).

[3] Sonntag. D., Engel; R., Herzog, G., Pfalzgraf, A., Pfleger, N., Romanelli, M., Reithinger, N.: SmartWeb Handheld - Multimodal Interaction with Ontological Knowledge Bases and Semantic Web Services. In: LNAI Special Volume on Human Computing, 4451, Springer, pp. 272-295, (2007).

[4] Wu, H., Grégoire, J., Mrass, E., Fung, C., and Haslani, F.: MoTaskit: a personal task-centric tool for service accesses from mobile phones. In: Proceedings of MobMid, (2008).