

Enhancing Internet Banking through Voice Control

Wolfgang Pree

C. Doppler Laboratory for Software Research
University of Constance

D-78457 Constance, Germany

Voice: +49 (0)7531 88-44 33 Fax: +49 (0)7531 88-35 77

pree@acm.org

Abstract. The paper describes how voice control was prototypically added to an existing Internet banking solution. It first sketches the conventional application and the aims that guided its extension. These goals influenced the design of the user interface. Thus the paper goes on to discuss the prevalent design issues of voice-controlled human-computer interaction in this context. A description of the prototype's implementation issues reveals insights how to integrate reusable components for voice recognition to a purely Java-based application. Visions of voice-controlled electronic commerce applications and plans for future work round out the paper.

Keywords: electronic commerce, voice controlled applications, human computer interaction

1 Project context

The project is carried out in cooperation with RACON-Linz Software GmbH (short RACON), a software company of the Austrian Raiffeisen banking group. RACON implemented Internet banking solutions as Java applets and made them available mid of 1997. A principal design goal is to address a broad range of customers, not just Internet freaks. As a consequence, RACON limited the features to a minimum set, mainly comprising transfers between accounts of domestic banks and European banks, as well as checking the balance of a customer's accounts. As only few customers do online stock trading, this functionality was separated into another applet. We refer to the core Internet banking applet, i.e., the one without stock trading, to ELBA-Internet, for Electronci Banking via Internet. The voice control interface was added to ELBA-Internet.

Due to the limited set of features, the user interface design of ELBA-Internet became straight-forward. The following screen shots illustrate the core functions of ELBA-Internet. The screen shots show the authentic German

labeling. The description of the screen shots provides a translation of the relevant terms. The next section focuses on the human-computer interaction design, describing how customers invoke these functions via voice commands.

The core application without voice control

After loading the applet and logging in, ELBA-Internet shows a list of accessible accounts and the current balance of these accounts (see Figure 1). A user can select one of the accounts and do one of the two core actions:

- To transfer an amount from the selected account, the user presses the button menu "Überweisung". He or she can then choose between a domestic transfer ("Inlandsüberweisung") or a transfer to a European bank ("Eurotransfer"). ELBA-Internet stores a few recent transfers as templates ("Überweisungsarchiv") from which the user can select.
- To take a look at bank-statement-like details, the user presses the button labeled "Umsätze".

The button "PIN ändern" is independent of the accounts and means changing the Personal Identification Number which is used for logging in to ELBA-Internet. "Beenden" is the German label for exiting, "Zurück" the label for going back one level.

Figure 2 shows the dialog for specifying a domestic transfer ("Überweisung"). After filling in the required account information, the user presses either the button labeled "Normal", meaning as soon as possible, or "Termin", where a date has to be specified for the transfer. (These buttons are not shown in the screen shot, as the "Normal" button was already pressed.) After specifying the type of transfer, the user has to enter a so-called TAN, short for transaction number. The TAN is a security feature. Bank customers receive a punch of TANs via snail mail on paper. For each transaction a customer has to provide a TAN.

Konten			
VerfügerIn: Bert Tester			
BLZ	Kontonummer	Kontozeichnung	Kontostand 05.04.1999
30025	1-23.456.783	Bert Maiers Gehaltskonto	12.530,54 ATS
30025	12-34.567.890	Bert Maiers Kegelclub	1.230,30 ATS
30025	123-45.678.901	Bert und Clara Maier (Familienkonto)	7.355,20 ATS
30025	123.455	Familie Maiers EURO-Konto	2.104,30 EUR
30025	554.321	Familie Maiers DEM-Konto	2.845,10 DEM

Wählen Sie bitte eines Ihrer Konten und die gewünschte Funktion aus


Raiffeisen. Die Bank 

Figure 1 Overview of accounts (“Konten”) in ELBA-Internet.

Figure 3 illustrates how ELBA-Internet displays bank-statement-like details. Besides scrolling the list the user can print (“Drucken”) these details.

2 Human-computer interaction issues related to adding voice control

Adding voice recognition to ELBA-Internet mainly pursues the goal of getting rid of the keyboard. Several

Web-enabled gadgets simply won’t have a keyboard attached to them. Examples are TV sets for Web browsing, probably mounted on the refrigerator, wrist watches, and mobile phones. Nevertheless, ELBA-Internet should feel and, if possible, look familiar to users of the PC version. For a detailed discussion of various aspects related to ‘talking and listening’ to computers we refer to Mountford and Gaver [2].

Überweisung			
Konto: 30025 / 123.455 EUR, Familie Maiers EURO-Konto			
Kontonummer EmpfängerIn	BLZ	Auftragswährung	Betrag EUR
2.114.726 <input type="button" value="Suchen"/>	34000	<input type="radio"/> ATS <input checked="" type="radio"/> EUR	85,00
EmpfängerIn (Name, Adresse)		Umgerechnet in ATS 1.173,37	
Union Yachtclub Attersee (UYCAS)		Bank EmpfängerIn	
A-4360 Attersee		RLB Oberösterreich	
AuftraggeberIn		Verwendungszweck	
30025 / 123.455		Regatta	
Familie Maiers EURO-Konto		Kundendaten (für Kodierzeile)	
000000000000< 00002114726+ 00034000> 000000008500<43+			
<input type="button" value="Beenden"/>	<input type="button" value="Hilfe"/>	<input type="button" value="Zurück"/>	TAN <input type="text" value="****"/> <input type="button" value="Abschicken"/>

Unterschreiben Sie bitte die Überweisung mit einer TAN aus Ihrer TAN-Liste


Raiffeisen. Die Bank 

Figure 2 Sample transfer of an amount to the specified account.

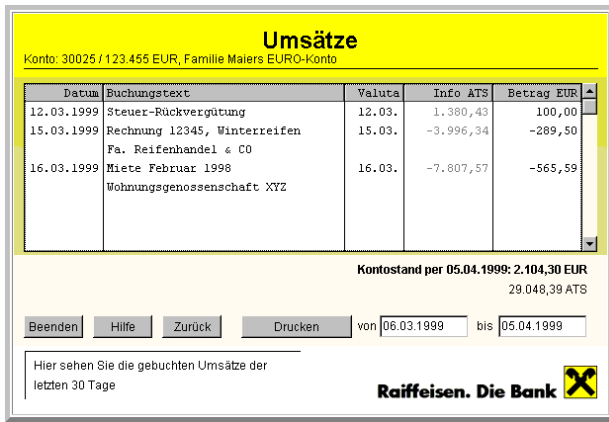


Figure 3 Account details.

For a user, remembering the words that are valid at a specific point represents a burden associated with voice control in general. The prototype of ELBA-Internet-Voice applies a simple rule for overcoming this problem: A small text box in the lower right corner of the applet displays all the words which make sense in a particular mode of the banking application.

Figure 4 shows the screen shot of the account overview with this additional text box. The user selects one of the accounts by telling the application the appropriate number. The recognized word is highlighted in the text box. The labels of the action buttons are also valid voice commands. As these words are visible in the user interface anyway, they are not replicated in the text box.

Voice-control hits its limits when information has to be entered into dialogs like the one for specifying a transfer. We think that the most appropriate way in the context of ELBA-Internet is that the user selects from a set of transfer templates. For example, a user might define templates for transferring money to four destination accounts, labeled Adele, Arnold, Steve and Yacht-Club. The idea is that a user or some bank employee assisting her, defines the templates for ELBA-Internet-Voice on a device with a keyboard, such as the PC, before using the voice-controlled application. The user of ELBA-Internet-Voice then gets these choices displayed in the text box. If he or she then speaks “Yachtclub” in our sample scenario, the information associated with the account is displayed in the edit fields (see Figure 5) so that only the amount (“Betrag”) has to be entered to finish a transaction. For the prototype version of ELBA-Internet-Voice an amount has to be spoken digit by digit.

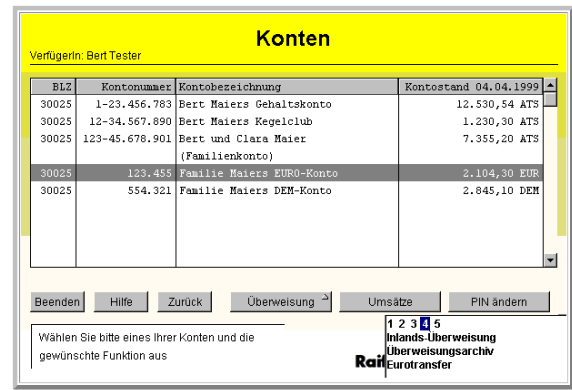


Figure 4 Showing valid voice commands in a text box.

If a user has to fill out the edit fields, i.e., if no template is available, the ELBA-Internet-Voice prototype requires spelling out the words letter by letter, which is too tedious. ELBA-Internet-Voice does not support free speech recognition, because keeping the application small represents an important constraint. Nevertheless, the lack of free speech recognition turned out to be no severe restriction, as users do most transactions based on the available templates.

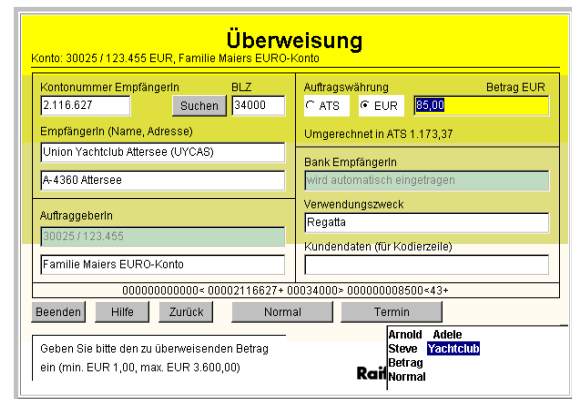


Figure 5 Interaction with dialogs based on templates.

ELBA-Internet-Voice avoids the introduction of modes wherever possible, for example, command and editing modes. An exception is entering the amount which represents a separate mode. The reason for this is that ELBA-Internet-Voice is optimized for the default usage: specifying the template, entering the amount and initiating the transaction. As soon as ELBA-Internet-Voice recognizes a template, the user enters the amount editing mode by saying “Betrag” (amount). Then the

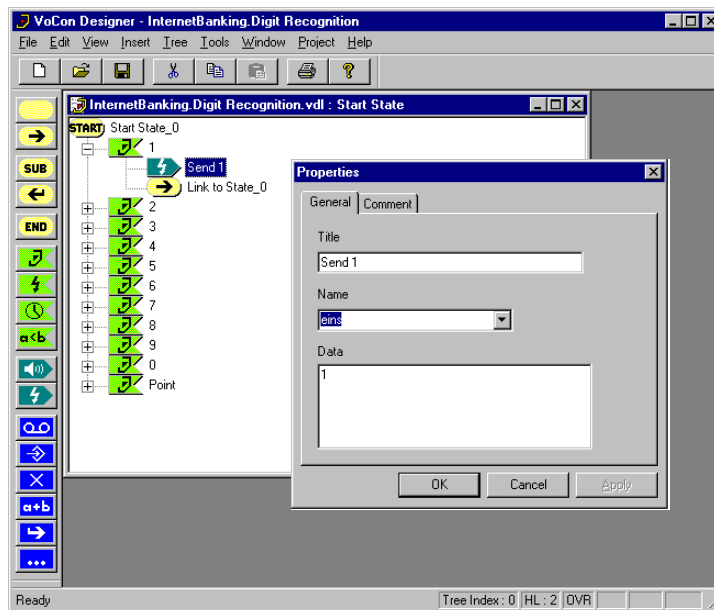


Figure 6 Sample state machine for recognizing single digits.

digits are spoken and the transfer type is specified by saying “Normal” or “Termin”. These commands are again at the same level as the commands for entering the amount.

Of course, TANs represent a hurdle in a voice-controlled Internet banking application. Imagine users away from their desktop or home who have to carry TANs on a sheet of paper with them. ELBA-Internet-Voice simply does not require the entering of TANs. Depending on the gadget where ELBA-Internet-Voice is available, TANs could be replaced by the user’s voice or finger prints.

The display of account details (see figure 3) is straight forward to control via voice commands and thus not further discussed.

3 Reuse and integration of a voice recognition component

For recognizing speech, ELBA-Internet-Voice reuses a Philips software component. This component is integrated with the ELBA-Internet Java applet. The Philips component was selected, because it is

- small in size (less than 100 KB for recognizing around 30 words trained by a specific user)
- well supported in a PC environment
- transferable to various Digital Signal Processors (DSPs)
- available

Philips provides an environment called VoCon designer to visually and interactively specify a state machine as basis of the recognition component [1]. Figure 6 shows the relevant parts of a state machine for recognizing single digits: For example, if the voice recognition component recognizes the word “Eins” (one), the string “Eins” should be generated and sent to a component connected to the recognition component. Afterwards the state machine returns to the initial state. Based on the state machine specification, the VoCon designer environment generates a resource file that is used by speech recognition component.

The connection between the Philips speech recognition component and another component is established in a PC environment via Dynamic Data Exchange, i.e., by means of a call-back function. The Philips speech recognition component invokes the call-back function of the connected component/application every time it recognizes one of the words specified in the state-machine. It sends the string associated with the particular word as parameter.

Thus the only precondition to reuse the Philips component was to integrate a Java component into ELBA-Internet that supports Dynamic Data Exchange. A switch-case-statement in the call-back function triggers the corresponding actions in the ELBA-Internet applet. Figure 7 schematically illustrates the coupling between the speech recognition component and the ELBA-Internet applet, both together forming ELBA-Internet-Voice.

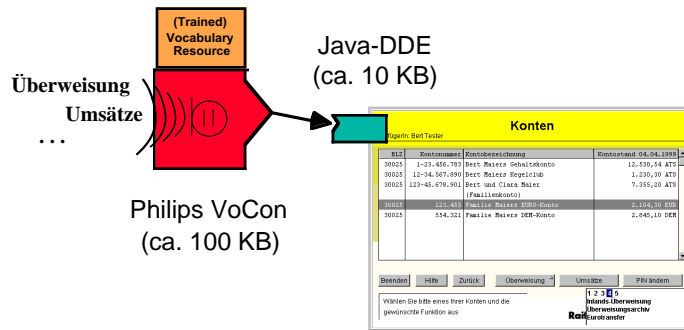


Figure 7 Integration of the Philips speech recognition component.

The VoCon Designer already generates a generic resource file for recognizing the words independent of a particular speaker. The speaker-independent generation is based on the writing of the words and works for German and English so far. In addition to this possibility, the words can be trained individually which further improves the recognition quality. The VoCon Trainer offers the interface to accomplish the training of the words listed in the specified state machine (see Figure 8). Basically, a user speaks all the words a couple of times. The spoken words are stored in wav-files. The VoCon Trainer finally constructs the speaker-dependent resource file for the recognition component out of the wav-files.

4 Visions for electronic commerce and future work

The usability engineering done with the ELBA-Internet-Voice prototype corroborates that the described enhancements of an electronic banking application can indeed make the usage of a keyboard superfluous. We view Internet banking a just one area of electronic commerce that might benefit from keyboardless input. As numerous Web-enabled gadgets without keyboards will appear on the market in the next years, keyboardless input could become an important factor for working with these devices.

The ELBA-Internet-Voice prototype also reveals some deficiencies of voice control. For example, words spoken by people who are around the user could trigger actions. Furthermore, the human-computer interaction in the realm of voice-controlled applications needs some standards similar to the way people interact with graphic user interfaces today. Extensive experiments with voice-controlled applications might lead to appropriate human-computer interaction standards. Nevertheless, gestures [3] might form an alternative to voice control. Thus, we will

investigate the application of gestures instead or in addition to voice control.

A short-term effort comprises the implementation of ELBA-Internet-Voice on a DSP, independent of a specific operating system.

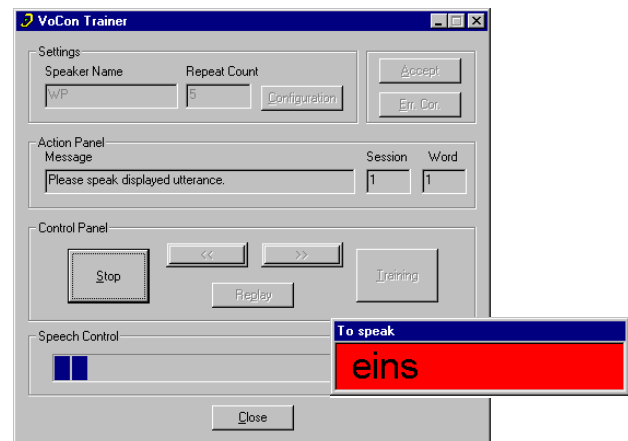


Figure 8 Speaker-dependent training of words.

References

- [1] R Groß (1999) *VoCon Designer User Guide*, Philips Speech Processing Division.
- [2] S Mountford and W Gaver (1990) Talking and Listening to Computers; in *The Art of Human Computer Interface Design* (ed. B. Laurel), Reading, Massachusetts, Addison Wesley.
- [3] G Kurtenbach and E Hulteen (1990) Gestures in Human-Computer Communication; in *The Art of Human Computer Interface Design* (ed. B. Laurel), Reading, Massachusetts, Addison Wesley.