A RECONFIGURABLE FORM GENERATOR TO CAPTURE DATA BY SPEECH ON MOBILE DEVICES

Ulrich Bieker

PRO DV Software AG
ulrich.bieker@prodv.de

Abstract: Why a reconfigurable form generator to capture data by speech on mobile devices? Currently, the development of voice recognition systems for mobile devices is time-consuming and expensive. In most cases, an individual solution must be implemented. We present a building block solution in order to develop applications to capture data by speech on mobiles devices efficiently. A reconfigurable form generator allows the adaptation and configuration of a simple application which is generated for a mobile device. A Personal Digital Assistant (PDA) and Windows CE is used as mobile device. The reconfigurable form generator is running on a desktop computer and generates the application and the voice user interface for the PDA. The generated application is intended to capture data and to issue commands supported by speech recognition. An important factor is usability and ergonomics of the system in order to allow end users without expert knowledge on speech recognition to build an application. This contribution describes industrial results of the EU funded project AMI-4-SME in the context of ambient intelligence.

1 Introduction

In order to allow end users a fast configuration of an application to capture data on mobile devices by speech, we propose a reconfigurable form generator as solution. E.g, if we consider the Hidden Markov Model Toolkit (HTK) [2], we can not expect that these tools are used by potential end users without expert knowledge on speech recognition. In order to enable a wide spreading of speech recognition applications, easy to use and ergonomic applications are necessary [3, 4]. The challenge is to implement an easy to build and reconfigurable speech recognition system.

In the context of mobile devices without wireless communication to a server, the solution must consider additional constraints like restricted memory and restricted computing power.

Figure 1 shows the general idea of a reconfigurable form generator. The intended application to capture data by speech is configured and adapted on a desktop computer. A dialogue manager will be used to define a possible dialogue of the user with the application. It shall support the definition of questions to the user and shall support the definition of the envisaged possible answers. Finally, the form generator will produce a grammar and a configuration file. By this, an application is generated, able to run e.g. on a PDA, enabling an end user to capture data on the PDA by speech.

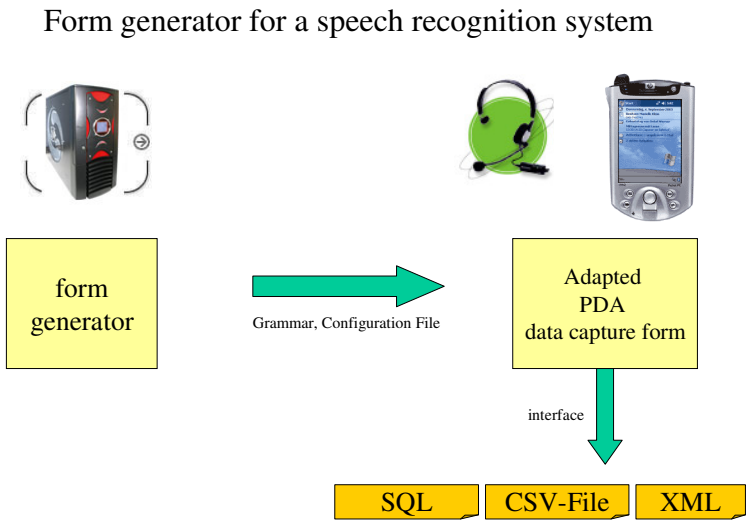

Figure 1 - Reconfigurable form generator to capture data on mobile devices

As presented in figure 2, the form generator shall also support the realisation of a data transfer from the PDA to a subsequent main intelligent system.

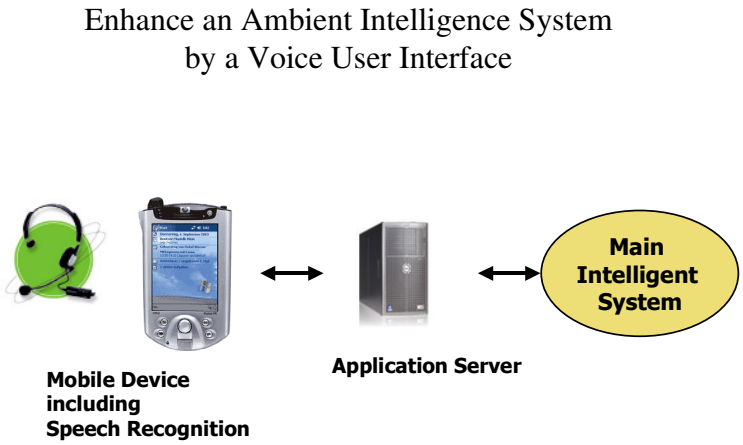

Figure 2 - Integration of mobile device and main intelligent system

Following features and advantages can be summarized concerning the reconfigurable form generator implemented as building block:

Features:

1. Generation of a simple and mobile device oriented application to capture data.
2. Generation of a robust and easy to use speech recognition application for a mobile device.
3. GUI based dialogue manager, reconfigurable with respect to heterogeneous business domains.
4. Realisation of a reconfigurable voice user interface.

Advantages:

1. easy to use
2. reusable software components
3. reconfigurable, customisable
4. adaptable
5. cheap

2 A reconfigurable Form Generator as Building Block

2.1 Key Functionalities

The reconfigurable form generator allows the adaption and configuration of a simple application for a PDA. The application is intended to capture data and to issue commands supported by speech recognition. The language (e.g. English, Spanish, German) of the generated simple application can be selected. The reconfigurable form generator may consist of two blocks: A reconfigurable dialogue manager and a reconfigurable voice user interface. PRO DV implemented the described approach as building block to be used in an ambient intelligence environment.

A desktop application "SRS form generator" is used to configure the components of a data capturing application on a mobile device (e.g. Personal Digital Assistant (PDA)). The desktop application is based on a database and offers a graphical user interface to ease the task of the configuration.

The SRS form generator is a flexible tool to generate different questionnaires with a vocabulary up to 50000 words. It allows the user to efficiently reconfigure a simple graphical (GUI) and/or voice user interface (VUI) for a PDA. Instead of implementing a fix application for a PDA it provides a set of reconfigurable input fields and instructions.

The SRS form generator is able to generate configuration files and grammars for the SRS of the mobile device.

All data captured on the mobile device by speech can be transferred by an interface, e.g. as SQL statements, as CSV or as XML file.

Figure 3 shows the main screen form of the form generator (running on a desktop computer) which enables the user to configure an application for a mobile device. The first version of the reconfigurable form generator consists of about 6 screen forms.

On the left side of the GUI, an outliner shows the hierarchical representation (model) of the application(s) to be generated. The user can select from different applications under construction. One example is an forestry application (timber inventory), i.e. to capture data about cutted trees. the application is able to recognize different attributes like tree species (e.g. oak, beech, …), wood damage (e.g. fungal decay, insects, …) diameter (in centimeter) or

length (in meter) of the trees. A vocabulary is assigned to each input field, e.g. the user can assign all tree types to the corresponding input field. Furthermore, a text to speech response can be assigned, i.e. the user can define the reaction of the generated application.

Details of an input field are shown on the right side of figure 3. The user can define the position of the input field on the screen of the PDA and the voice recognition command to select the input field. Furthermore, the data type of the input field (like string, number, date or boolean) can be defined. Data length and additional attributes of input fields can be defined: required input field, optional input field, enable or disable fields. Finally, a speaker adaption might be possible by loading a predefined speaker profile. However, as far as our experience goes, a speaker adaption for such a mobile application is usually not necessary.

After having generated and transferred the application, the data capturing form can be used on a PDA. For example, the user can tell the system the name of the input field to be filled, e.g. “tree species”. Afterwards, the user speaks the input, e.g. “oak”. Like this, all input fields of the application can be filled. Finally, the user speaks “save”, to save the captured data. For each input field, the system has a catalogue of allowed expressions. E.g., the field damage classification knows about 15 different types of damages like insect attack, mycosis etc.

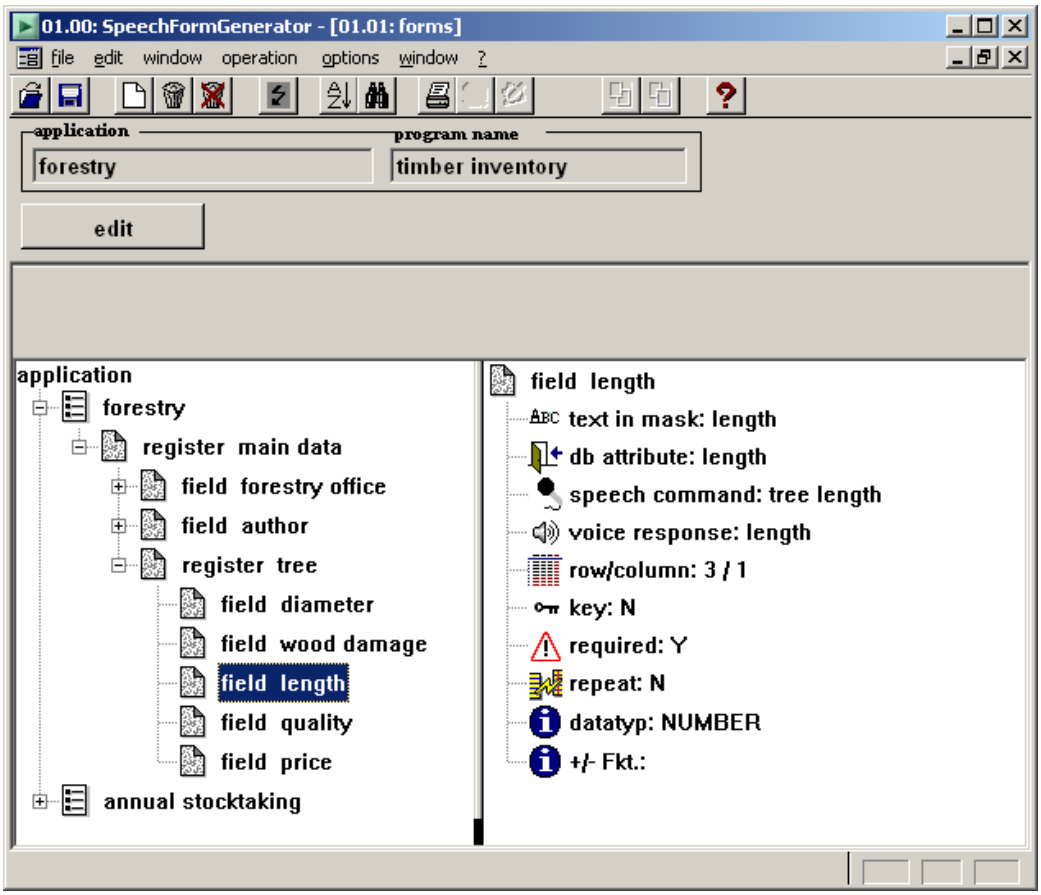

Figure 3 - Example screen shot of the reconfigurable form generator

2.2 The generated mobile application

The generated application itself can be controlled by a voice user interface. A classical graphic user interface is available as well.

Reconfigurable object types of the mobile application might be:

1. Window forms
2. Buttons / commands
3. (Unique) Input fields
4. Output values, answers to the user
5. Interface

Each above mentioned object type has a set of defined actions which can be controlled by speech:

1. Window forms can be opened or closed.
2. A set of known commands can be used like: *save data, export, start, stop, …*
3. A set of input fields can be configured as follows:
    a. Data type: string, number, date, boolean
    b. Data length
    c. Attributes of input fields: required input field, optional input field, enable or disable fields
    d. Allowed input values – catalogue of excepted spoken words per field. Optional: choice modus.
4. Output values can be spoken by the TTS - the text to speech engine, which is included as well. A set of standard-error-messages may be defined.
5. Reconfigurable output interface e.g. as CSV file, SQL-script or as XML file

3 Examples of Use

Data capturing by speech recognition on a mobile device can be useful in many examples of use:

1. Inventory: In many companies, a regular inventory is necessary. E.g., it is possible to capture data like item code, quantity and status by speech recognition.
2. Capturing data in the field, e.g. in forestry or agriculture.
3. Using GPS on a mobile device, a lot of use cases for mobile processes - e.g. field services - can be supported by speech recognition. The integration of GPS, mapping applications (GIS) and speech recognition allows a huge amount of ambient intelligent scenarios, specifically by realising end-user task support which is tailored to the specific situation, context and location.
4. Controlling and reconfiguration of manufacturing and assembly lines (AMI-4-SME [1]).
5. Accounting of activity recordings from maintenance engineers: E.g., capturing data like maintenance time in hours, driving hours, customer, date of maintenance etc.

4 Technology and Development Environment

In the following, we describe in detail which hardware and software is used. This aspect is important in order to achieve a good speech recognition rate.

4.1 Software

As basic technology - in order to enable voice recognition and audio response - the software VoCon 3200 and RealSpeakSolo from Nuance [5] is used.

4.2 Desktop Computer

The application on the desktop computer has been developed with Gupta SQL Windows. Windows XP is used as operating system. The data are stored in a SQL-Base or ORACLE database.

4.3 Mobile Device

The application on the PDA has been developed with Microsoft Embedded Visual C++. An IPAQ H5550 of the company Hewlett-Packard is used. It possesses a 400 MHz ARMV4 processor, 128 MB storage area and a Pocket PC 2003 operating system.

4.4 Headset

Actually, we use the Sennheiser Headset CC 550.

5 Results

The analysis of SME innovation needs in manufacturing industry and realisation of a concept prototype supporting an SoA based ICT building block infrastructure and architecture in the AMI-4-SME project served as input for further experience prototyping with additional PRO DV customers from other business domains (i.e. government and forestry). Hence, a first kernel system of the presented reconfigurable form generator was specified, implemented and tested with the ministry of Baden Württemberg in Germany and a private forestry company in Germany. Currently, both customers evaluate the application and first users are experiencing the generated voice user interface already in the field.

The algorithm implemented in the VoCon 3200 engine considers several characteristics of the incoming signal to decide if the input signal contains speech or not. A minimum speech duration constraint avoids that the engine reacts on short high-energy events such as clicks or door slams. An absolute energy threshold can be set to avoid that the engine triggers on very low energy events. Even if these low energy events are speech events, it is probably not a good idea to trigger on them because they are most likely not the speech the recognizer is interested in or the system gain is not correctly set up. The absolute threshold range is from -72dB to 18dB. In our application an absolute threshold range of -35 dB is used. Hereby, the speech recognition rate (without speaker adaptation) is between 95% and 99%. Using the headset mentioned above we achieve the best results (up to 99%) and a background noise up to -42 dB is acceptable. Beyond -42 dB, the speech recognition rate decreases drastically.

In order to use a wireless Bluetooth headset, it is not easy to find an adequate headset and PDA to achieve good results. It is necessary to use the Bluetooth Audio Profile, i.e. the ACL profile Advanced Audio Distribution (A2DP-SNK as sink and A2DP-SRC as origin). However, actually it is not easy to find a headset supporting the A2DP-SRC profile and to find a mobile device supporting the A2DP-SNK profile and Windows CE.

Based on these first results, within the AMI-4-SME project, PRO DV will further enhance the form generator as a widely applicable, portable and reusable building block in the scope of a Service oriented Architecture concept, aiming at an integration of the different building blocks from the ICT vendor partners in the AMI-4-SME consortium (i.e. Telefónica and Softronica). Therefore, further extensions and adoption of additional technology concepts must be designed and implemented.

Gelöscht: ¶

Formatiert: Nummerierung und Aufzählungszeichen

6 Summary

One advantage of our approach and the resulting application for a mobile device is: A continuous wireless communication to a server is not necessary - a big advantage in the field, e.g. in areas with dead spots in the forestry. Furthermore, the reconfigurable form generator is a tool which enables end users without expert knowledge on speech recognition to build their own application. Our solution enables end users and customers to become familiar with applications supporting speech recognition.

Acknowledgements

The approach presented describes industrial results of the EU funded RTD project AMI-4-SME: Ambient Intelligence Technology for Systemic Innovation in Manufactoring SMEs [1] (www.ami4sme.org), supported by the Commission of European Community, under Information Society Technology Programme.

List of References

[1] AMI-4-SME, Revolution in Industrial Environment: Ambient Intelligence Technology for Systemic Innovation in Manufacturing SMEs. http://ami4sme.org, EU IST Project Contract Number 017120. October 2005 to September 2008

[2] HTK Speech Recognition Toolkit, web pages and download site at htk.eng.cam.ac.uk

[3] Cohen, Michael H.; Giangola, James P.; Balogh, Jennifer: Voice User Interface Design.Addison-Wesley, 2004.

[4] Makhoul, John; Schwartz, Richard: Human-machine communication by voice. 1994, National Academies Press.

[5] Nuance Communications, www.nuance.com