Voice User Interfaces in Industrial Environments

Silke Lotterbach, Matthias Peissner

Competence Centre Human-Computer Interaction
Fraunhofer Institute for Industrial Engineering (IAO)
Nobelstr. 12
70569 Stuttgart
Silke@Lotterbach.de
Matthias.Peissner@iao.fhg.de

Abstract: Voice user interfaces (VUIs) have become very common for telephone based applications, e.g. call centers and value-added services. Though there are advantages for further fields of application as well, they are still rarely used. This paper describes the range of aspects to be considered before implementing VUIs in industrial environments. It supports the decision process on whether or not a VUI is reasonable for a certain context of use. In order to create an overview of the relevant aspects, this paper refers to findings from the literature and integrates insights gained from interviews with carriers as well as suppliers of VUI applications.

1 Introduction

Voice user interfaces (VUIs) use speech technology to provide users with access to information, allow them to perform transactions, and support communications [CO04]. VUIs rely on automatic speech recognition (ASR) for user input and use speech, either recorded or synthesized, as their primary form of output. VUIs enable human-computer dialogues that cover only a limited spectrum of human conversation focusing on a predefined set of speech acts in a specified task domain. These restrictions reflect the limitations of today's speech recognition technology.

Today, speech technology is widely used in call centre applications. In other domains, however, speech interaction is currently only rarely used. Especially in the field of industrial production, the application of speech technology has still a great potential for growth. Operators with radio-controlled headsets benefit from higher mobility as the input and output of information is not restricted to a fixed location of a computer terminal. An employee of a saw mill, for example, can adjust the breadth of the saw without having to put down his work piece and to take off his gloves. This example shows that in the industrial domain the requirements of the VUI are often even less complex than in other domains. Here, the context of use is well defined and often "known" by the system. Moreover, the user interface is controlled by a small number of recurrent users, which means that both system and user can be trained for an efficient interaction.

This paper discusses the conditions for a beneficial application of VUIs in industrial contexts. We describe advantages and disadvantages of speech interaction in order to point out which contextual conditions have to be considered before employing VUIs.

2 Criteria for the implementation of voice user interfaces

2.1 Communication with computers is still restricted – yet effective

Natural language has a huge advantage compared to fixed menu structures: users can compile a lot of information in short sentences, that otherwise would be hidden in deep hierarchic structures. The short sentence „I want to fly from Stuttgart to London tomorrow morning“, contains the number of people traveling, date, time, start and destination of the journey. Additionally, the user can describe in own terms what he wants, rather than adapt to the systems language. VUIs can – in contrary to graphical user interfaces (GUIs) – process different expressions for a command (e.g. show, visualize, view or display). Today, communication with technology is task oriented and therefore restricted, but if applications are well designed, VUIs offer a direct and efficient way of interaction.

2.2 Even large amounts of information can be easily input via voice, whereas output is limited

Most users can speak significantly faster than they can type or write [ZU99]. Speech input can save time and avoid problems. Listening to speech output – on the other hand – is slower than accessing visual information. People can read faster than they can listen. Also, people can easily compare texts, pictures or figures visually. Doing this via a VUI is more demanding while users need to keep a lot of information in short time memory, which is not comfortable and may lead to mistakes. Suitable tasks for VUIs can require high amounts of user input, whereas the system output should be either an action or rather small amounts of information at a time.

Option menus can be displayed on GUIs very easily, but GUIs can only display a certain amount of information and displays are overloaded regularly. The options available for a VUI are often not visible and therefore it is hard for users to know what systems can perform. It is important that users either know their capability or additionally have a visual aid (e.g. screen) to display possible options if these are too complex to be output via voice. The access to information is easy as long as the user knows it is accessible.

Acoustic output allows information to be exposed to all users at a time. Acoustic output is widely used for emergency signals, and it can also be used to reveal which action a machine is performing etc. (e.g. lifting platform moves down). Users are enabled to react immediately if attendance is necessary. Without mobile devices, the operator notices errors only when passing the affected machine by chance or when he gets back to his PC the next time. The opportunity of receiving real-time messages anywhere and anytime enables the operator to immediately react to critical events (cf. [JA04]).

2.3 Dealing with linguistic diversity is practicable, some kind of information is less suitable

VUIs can be trained to understand different languages as well as accents or dialects. This may lead to a unique role of VUIs in the European Union due to cultural and linguistic diversity.

Spatial and visual information is difficult to be transferred into speech (e.g. maps; plans; commands like shift, turn; descriptions of motion as in sports) and is therefore hardly suitable for VUIs. The same applies to emotional information, as it is not only transmitted by words, but also by intonation, gesture and facial expression. Another problem is the acoustic presentation of structural information. Furthermore, it is not possible to present information in a parallel way, as is possible with visual information (e.g. tables). Users can quickly switch between two pictures or pictures and text, but switching between two blocks of acoustic information is difficult. Also the manipulation or correction of visual data is possible, whereas the same action applied to acoustic information is time consuming and complicated, e.g. spelling mistakes (cf. [SH00]).

Suitable information for VUIs is such that it can easily be put in words [GA99]. Fixed values for example are easily transmittable via voice. Security relevant information is suitable when user identification is required. It is not suitable if critical information is transferred (others may listen) or for critical user input such as an emergency stop (as the recognition rate is not 100 % and therefore users should not rely on it in the first place).

2.4 Overcoming physical handicaps

For users with tendovaginitis, motor impairments or just big hands it is often difficult to operate (mobile) computers manually. The implementation of a VUI can simplify this kind of work significantly. Also users with visual impairments can profit from VUIs, whereas users with aural/vocal impairments can interact via voice only in a limited way.

2.5 Hands and eyes free, mobility provided

VUIs can be controlled via wireless headsets. This enables location-independent human-computer interaction. Users can walk around freely and take a comfortable posture, which helps to prevent postural deformities. Wearable VUIs economize not only the routes between work piece and PC terminal, but also the paths that the user's eyes and input devices (e.g. mouse) have to take on the screen.

Speech applications provide a particular advantage when the user has to pay full visual attention to the primary work task while interacting with the system. The same holds for situations when the user's hands are busy or not free for other reasons, e.g., wearing gloves, being soiled. In other words, conventional GUIs that demand visual attention and manual control interfere with most primary work tasks of industrial environments.

VUIs support efficiency by making it possible to accomplish two tasks simultaneously, e.g., operate a measuring device and input the displayed values into the system. This is not only beneficial in terms of time savings, but it is also supporting an ergonomic flow of work which is not disturbed by secondary tasks (e.g. keyboard entry) [KOE03]. Thus, VUIs can improve productivity and precision [CO93].

2.6 Straight usability – even under adverse conditions

Acoustical systems have great advantages if the perception of visual information is impaired (e.g. darkness, too much light, vibration). Situations like this occur for example during the maintenance of machines, when flaps or covers conceal sources of light.

VUIs provide special advantages "for work under uncomfortable circumstances – humidity, low temperatures, insufficient lighting, e.g., in deep freeze stores or during inspections outdoors" [HE04]. The only equipment to be worn is a headset and robust radio receiver. In particularly hygienic environments a headset can easily be kept clean. VUIs are also beneficial, when there is no or only a very restricted keyboard or screen available (e.g. mobile phones, mobile devices).

Today's speech technology can handle noisy environments relatively well. Especially continuous noise is easily manageable, whereas unforeseeable and sudden peaks may lead to difficulties. In contrast, the implementation of VUIs may not be favorable in very quiet environments or whenever human to human communication is vital.

2.7 Motivational aspects matter

Users may be not too happy at first, when VUIs are introduced into their working environment. Negative attitudes are more widely spread than positive usage experiences. Only a highly acceptable VUI will be able to create user acceptance and to utilize the advantages of VUIs, e.g.:

- Efficiency can be improved by omitting unpleasant and time-consuming tasks.
- Computer knowledge is not necessary for using a VUI.
- Users are often used to wearing earplugs and therefore accept wearing headsets very quickly, as this gives them the opportunity to communicate.
- VUIs can be adapted to individuals in a very short time. Therefore, recurrent users can choose the terms/language they want to use and do not have to avoid foreign words or accents and dialects.

3 Conclusion

VUIs will not be powerful enough to replace graphical user interfaces in the future. According to Shneiderman, "speech is the bicycle of user-interface design: it is great fun to use and has an important role but it can carry only a light load" [SH98]. However, under certain conditions and in certain contexts, speech – the same holds for bicycles – is the most comfortable and most efficient way to get things done.

In this paper, we have analyzed the most important factors that influence the beneficial employment of VUIs in industrial environments. Further studies and interviews with stakeholders are planned in order to see how the specific requirements in individual production contexts can be met by a selective and purposeful application of speech technology. The potentials of speech interaction presented here make us expect significant market growths for VUIs in the production context. Special attention will be paid to multimodal interfaces that combine the advantages of auditory and visual-manual interaction. According to the specific requirements of the current task and working context, the user can switch between interaction modalities or use a suitable combination of them.

In order to tap the full potential of these future user interfaces and in order to achieve a broad acceptance among the users and carriers of the emerging systems, a user-centered design approach will be indispensable.

References

[CO04] Cohen, M.; Giangola, J. P.; Balogh, J.: Voice User Interface Design. Addison Wesley. 2004.

[CO93] Cohen, P.R.; Oviatt, S.L.: The role of voice in human-machine communication. In: Voice Communication between Humans and Machines. Editors Roe, D.; Wilpon, J.: National Academy of Sciences Press, Washington, D. C. 1993.

[GA99] Gardner-Bonneau, D., Editor: Human Factors and Voice Interactive Systems. Kluwer Academic Publishers. 1999.

[HE04] Helbig, J.; Schindler, B.: Speech-Controlled Human Machine Interaction. it – Information Technologie. Issue 46 (6/2004). Oldenbourg Verlag. 2004.

[JA04] Janssen, D.; Peissner, Schlegel, T.: FabSCORE - A framework for ergonomics in semiconductor productions. In: Khalid, H.M.: Work with computing systems: Proceedings of the 7th international conference on WWCS, 29 June – 2 July 2004, Kuala Lumpur, Malaysia; bridging diversity at work Kuala Lumpur, Malaysia: Damai Sciences, 2004.

[KOE03] Köppe, C.: Betriebswirtschaftliche Evaluation von Spracherkennungssystemen. Wirtschaftswissenschaftliche Fakultät der Universität Hannover. 2003.

[MA96] Markowitz, J.A.: Using Speech Recognition. Prentice Hall PTR. 1996.

[SH98] Shneiderman, B.: Designing the user interface, third edition, Addison-Wesley. 1998.

[SH00] Shneiderman, Ben (2000) 'The Limits of Speech Recognition: To improve speech recognition applications, designers must understand acoustic memory and prosody', COMMUNICATIONS OF THE ACM, September 2000/Vol. 43, No. 9, p. 63-65

[ZU99] Zue, V.: Talking with Your Computer. Scientific American, August 1999, retrieved 10.1.2005 at: http://www.sciam.com/article.cfm?articleID=0009D2B7-F2E6-1C72-9B81809EC588EF21&catID=2. 1999.