Speech-Controlled Wearables: Are We There, Yet?

Christian Bürgy[1], Edwin Vogt[2] and Ilija Komljenovic[3]

Xybernaut Corporation, Research and Development,
Otto-Lilienthal-Strasse 36, D-71034 Boeblingen, Germany
http://www.xybernaut.de

[1] c.buergy@xybernaut.de
[2] e.vogt@xybernaut.de
[3] i.komljenovic@xybernaut.de

Abstract. Although using speech to operate a wearable computer system is probably the best hands-free interaction currently available, there are only few applications based on speech-controlled wearables. Both, hardware and speech recognition software improved during the last years, but still industry seems to hesitate with the introduction of wearable computers. This article presents an overview on the design of speech-controlled wearable computer systems and illustrates the development of such systems by comparing the results of a previously conducted project with current and up-coming technology.

1 Introduction

The term "applied wearable computing" implies truly usable wearable computing systems, which support users while they perform activities that need some kind of assistance by information technology. Wearable computers are hardware and software systems, worn on the body, and are supposed to be unobtrusive. To not to restrict the wearer / user in performing the actual activities, the interaction with the device has to stand back behind primary activities. Often times, the user needs both hands for a primary activity and thus, it is beneficial that the wearable computer offers hands-free user interfaces for man-machine interaction, such as speech recognition technology. Hands-free computing does not necessarily mean that the user has no option to operate the device using one or two hands, but implies hands-free use at least in some situations when the user needs both hands for doing the actual job [1].

The idea of speech-controlled wearable computers exists for years now [2][3], but despite pilot projects [4][5], there are few speech-controlled wearable computer systems applied to industrial applications – maybe with the exception of pick-by-voice systems [6][7]. In this article, we want to illustrate our path to speech-controlled wearables and the challenges in designing such systems that are ready to be applied in industrial applications. We believe it is necessary to think over the integration of speech technology and wearables to see how industrial and academic research can join efforts in enabling the market for speech-controlled wearables.

2 Background: Towards a Speech-Controlled Wearable

Hands-free computing is the unique selling proposition (USP) for speech-controlled wearables. The problem is that there is no one-size-fits-all speech-controlled wearable, yet. The challenges that we face in the design phase of each new project result from different hardware needs and especially from the still not trivial task of designing speech interfaces. In the following, we describe where we stand - in terms of hardware and software of speech-controlled wearables - and how we got there.

2.1 All-Purpose vs. Customized Wearables

Wearable computers took a dramatic change since the first (commercial) generations in the early Nineties. While processing power, a prerequisite for using speech technology, increased, the form factor was reduced – always with the compromise between performance and size.

Commercially available wearable computers at Xybernaut® took few generations (Figure 1) and offer general purpose products that are ready-to-use. These wearable computers offer standard interfaces, comparable to laptop computers, and can be enhanced by peripherals and accessories, such as head-worn displays (HWD), wireless communication means or wearable keyboards.

Fig. 1. Development of Xybernaut® Mobile Assistant® wearable computers

In relation to speech technology, the hardware development in that period could just keep up with the demands on processing power. Although more processing power can still increase recognition rates, speech engines recently could run with processors with less than the highest-available performance, which gives some more design space for configuring speech-controlled wearable computer systems and enables systems with lower power demands.

If we take the development of wearables in academic research we see a clear difference in the design approach compared to commercial development. For example the “Family Tree of CMU Wearable Computers” [8] shows over twenty design generations for the same time period as in Figure 1. These designs are task-specific wearables that were built for single projects and had functionality directed to the needs of the application at hand. Consequently, not all of these systems would be suitable to be speech-controlled, because they did not need any hands-free interfaces and thus were designed differently.

This comparison between commercial and academic research and development shows the challenges in designing wearable computer systems. These result, besides the technical issues and limitations, mainly from contradicting market demands. Either there are demands on high processing power, such as Augmented Reality (AR) or other graphic-intensive applications, or the system needs to be unobtrusive and with long battery discharge cycles, such as in inspection or retail applications.

2.2 Customizable Wearables

Since it is difficult to build an all-purpose standard wearable, there are some commercial specialized wearables [6][9] that occupy a niche and thus are quite successful in the emerging wearable computer market (Figure 2). Other examples show pilot applications and field tests with a laptop computer in a backpack. This is a legitimate workaround for applications which have higher performance needs than offered by commercially-available systems [10]. However, for meeting industrial standards, future (commercial) wearable computer systems should be customizable to a certain degree as to allow for satisfying a broader range of market demands. This customization does not necessarily mean that the system has to be adaptable at operation time, but that its hardware and software components can be adjusted to the application at hand.

Fig. 2. Vocollect Talkman® (left) and Symbol Technology Wearable Scanning System

In the example of the Xybernaut® Mobile Assistant®, this development can be seen as well: in the beginning, handheld flat-panel displays (FPD) were sufficient as they enabled high-performance computing in a wearable/portable form factor. To avoid the obtrusive cable between FPD and the CPU unit, Xybernaut® offers the Atigo® wireless flat-panel, which provides a wireless display with access to sufficient processing power. As sub-notebooks and TabletPCs enter the market and the Atigo® became more powerful, this concept is about to taper out in favor to one self-contained hand-held unit. As a result, there is a tendency towards two categories of speech-controlled wearables: the “classic” wearable either with an HWD or completely audio-based; and a speech-controlled handheld, which a user can operate by speech while wearing it in a pocket and which offers a flat-panel display when needed (as such, it functions as kind of a part-time wearable / part-time hand-held system).

3 Speech Interaction with Wearables

Besides the hardware challenges, the major issue of building speech-controlled wearables is to find the right support by speech technology [11], based on the requirements of the application. Only after defining the data the user has to enter or retrieve and the situation in which the data exchange occurs, the right speech interaction can be determined [12]. The following section illustrates the basic factors of speech technology with respect to the use on wearable computer systems.

3.1 Speech Support on Wearables

The first thing to determine while developing a speech-controlled wearable is in which situation hands-free interaction in form of speech interfaces is required and whether other (non-hands-free) user interfaces can be used as well. Thus, the speech interaction can be constantly available or available to support the user only in certain activities. Thus, we classify speech support for wearables in the following categories:

- **Audio-only:** no display needed, e.g. pick-by-voice systems;
- **Audio input / multimodal output:** data collection occurs while the hands of the user are blocked by the actual job, results are shown with a head- or body-worn display and augmented by audio feedback; and
- **Multimodal input and output:** user can use input by hand and speech simultaneously; results are shown with a head- or body-worn display and augmented by audio feedback.

All three categories will probably occur in the same application at different situations. But if the application allows for limiting to one of the first two categories, the system can be scaled down significantly.

3.2 Speech Interaction Modes

Another consideration in speech interaction design is to choose between different modes of speech recognition and automated speech output:

- **Command-based recognition (CBR):** recognition of single words or sets of words as predefined commands in discrete or continuous recognition (example: selection lists in telephone support hotlines);
- **Dictation-mode recognition (DMR):** recognition of "free text input", i.e. based on speech input, complete sentences and paragraphs are recognized, according to language models, grammars and specific contexts (example: dictation of business communication into word processing software);
- **Pre-recorded sounds (PRS):** audio feedback is stored as audio files and played as a whole or composed by text fractions (example: airport announcements);
- **Text-to-speech (TTS):** text is read by a synthetic voice from within the software application (example: speech-enabled web browsers and e-mail tools).

While recognition modes, i.e. input modes, vary in demands for processing power, the output modes set constraints on the flexibility of the application structure.

DMR demands high processing power. By restricting application to CBR, the workload drops and consequently discharge cycles of batteries extend. To date, even some mobile phones and PDAs reach satisfying results with CBR, whereas dictation is only about to become available on mobile platforms. One workaround is to use recorded audio files, which can be sent to a server for later post-processing.

PRS as output restricts the application output to a pre-defined vocabulary. While this mode might sound more natural, TTS is more flexible as it allows reading out any text – with restrictions for special terms and abbreviations.

3.3 Speech Interaction Design

Besides deciding, which interaction mode and which input / output interfaces to use, the developer of an application has to pay attention to the speech interaction design. This is especially true for industrial application, in which the use of the speech-controlled device is only a secondary task that supports the actual activity. This leads to some requirements on the application as illustrated by Helbig [13]:

- The application has to be robust and insensitive to environmental noise;
- Speech control has to be intuitive and flexible;
- Recognition errors should be treated by an intelligent correction mode; and
- If there is no visual feedback from the machine (in a given situation), the dialog should give the system state and the last input.

Furthermore, to be able to design the right interaction, it is important that the speech engine provides the functionality necessary for the application at hand and, while using it with wearables, that the engine fits the resources of the mobile platform.

4 Speech-controlled Wearables

As pointed out before, it is very difficult to define one speech-controlled wearable that fits all applications demands. In the following, we want to compare an earlier project based on Xybernaut® hardware with currently available systems and derive an outlook from that comparison.

4.1 Five Years Ago: Automotive Inspection

The goal of a research project conducted in 1999/2000 with CMU and Bosch [14] was to explore and demonstrate the usability of a speech-controlled mobile device that supports technicians during vehicle inspections in a garage environment. Most vehicle inspections follow a similar process:

- All incoming vehicles are registered and the vehicle data is entered in a central office or at the reception;
- The vehicle inspection process consists of a sight inspection that often comprises a high number of inspection items (in this project about 500) that, if faulty, have to be marked in a report;
- This list of inspection items is printed at a centrally located computer and taken to the pit for the actual inspection;
- The technician enters his or her findings into the computer and sends the results to the office computer where the report is printed; and
- Data from networked testing equipment is sent directly to the centrally located computer and included while generating the report.

Based on a requirements analysis, demands on the device that was intended to support the technicians were the following:

- The device has to be mobile, worn on a belt or carried in a pocket;
- It must be possible to operate the device completely hands-free, using speech control;
- The system must be functional in noisy and dirty environments;
- The device has to support all tasks to be performed during the targeted vehicle inspection; and
- Data to be exchanged has to be transmitted wirelessly to a centrally located computer.

According to these requirements, a pilot speech-controlled wearable computer system was developed consisting of a Xybernaut® Mobile Assitant® IV running a Learnout & Hauspie speech engine (Figure 3). Table 1 shows the system specification of the wearable inspection computer.

Fig. 3. Xybernaut® MA® IV used in a speech-controlled automotive inspection

Table 1. Key features of speech-controlled wearable for vehicle inspections

	Processing Power	**Battery Discharge Cycle**	**Physical Dimensions**	**Connectivity**	**Display**
Xybernaut Mobile Assistant IV running MS Windows 2000 and L&H speech engine	Intel Pentium 233 MMX; 160 MB RAM	about 1.5-2 hrs using external 9-cell battery	about 1200g including battery, without display; 190x41x119mm	2 x PCMCIA; USB; serial	proprietary flat panel or HMD

In field tests that were conducted in real garage environments and with garage technicians that usually inspect vehicles in the traditional way with a pen and a clipboard, we got valuable feedback on the different components of our demonstrator system:

- The hardware was felt too bulky, too big, and too expensive;
- The discharge cycles were felt to be too short;
- Some technicians complained about the futuristic appearance of an HWD;
- Technicians also disliked the usability of a head-worn display during inspections of vehicle areas with difficult accessibility, such as the wheel cases;
- The manner in which the software had to be navigated to reach a single inspection item using speech commands was new and caused some irritation for the technicians; and
- At certain times, the list of inspection items did not show completely on the screen and the users had to scroll the screen using speech commands.

In this field test, most complaints where directed to the hardware. Concerning the speech control, it was not the recognition rate, but the design of the speech navigation and how the related visual feedback was presented, which was criticized. This leads to the assumption that hardware improvements would directly result in improvements for the whole system.

4.2 Today: Currently Available Systems

Hardware developed significantly – mostly in display technology and in low-power processing units. Table 2 lists example configurations of currently available wearable computing systems. Some of the feedback of the technicians of five years ago still applies to these systems:

- Hardware prices went down and will drop even more when we enter a more mature market – depending on the target application, scaled-down systems are an option as well;
- Hardware became less bulky and easier to “wear”;
- There is still the trade-off between system size and battery discharge cycles, but systems are close to reaching the eight hours shift with one cycle – compared to around two hours in a similar form factor five years ago; and
- Head-worn displays became more usable, i.e. lighter, clearer and with higher resolution; consequently, the willingness to use wearables increased slightly.

Table 2. Key features of currently available commercial wearables (examples) [15][6]

	Processing Power	**Battery Discharge Cycle**	**Physical Dimensions**	**Connectivity**	**Display**
Xybernaut Mobile Assistant V; running MS Windows XP or Linux; different speech engines available	Intel Mob. Celeron, 500 MHz, 256 MB RAM	about 2-3 hrs using internal 3-cell battery; with extra battery in holster about 5-6 hrs	about 450g including battery, without display; 150x90x50mm	Compact Flash, USB, FireWire, proprietary display connector; With holster: PCMCIA, USB, FireWire, VGA	proprietary flat panel or MicroOptical SV-Viewers, up to SVGA
Nomad system, running MS Windows CE.net; speech control not yet available	Intel XScale 400 MHz	up to 8 hours	794g without display	2 x USB, 802.11b	SVGA retinal display, monochrome
Vocollect Talkman 2, running MS Win CE.net and proprietary speech engine	Intel StrongARM	up to 8 hrs with extended battery	about 500g	serial, 802.11b	no display - audio-only device

Commercially available speech technology has improved significantly in recent years and has made its way into industrial applications as well as into information systems and consumer products. We find speech recognition and synthesis in car navigation systems, service telephone hotlines, and mobile phones. All the above examples are using the CBR mode of speech recognition. For most applications, CBR is sufficient and thus, we see most value in targeting these applications with speech-controlled wearables.

Within the last five years, speech engines got more robust and gained functionality, such as improved barge-in, error-handling, voice recording or demo modes. Most of the robustness was reached by improving the algorithms and language models; the additional functionality helps to develop more user-friendly speech navigation for applications. Furthermore, the footprint of the engines is reduced so that they can be used on less powerful devices than before.

From the acceptance perspective, we can say that users get more and more familiar with speech technology and mobile devices as they become more widespread and integrated in daily life. Although operating an industrial tool, which the wearable will eventually be, is still a different task, people become more open to use mobile IT support.

Compared to five years ago, speech-controlled wearables got slimmer and more robust, but are not yet where the developers and the user of such system would like to see them. It is now more than ever in the interest of the "Speech-Controlled Wearable" that prospective users and system designers work together with the hardware developers to define the ideal system.

4.3 Tomorrow: The Ideal Speech-Controlled Wearable?

In an attempt to define the ideal "Speech-Controlled Wearable", we want to list requirements derived form earlier experience and today's feasibility of hardware design, to start a discussion or some kind of wish list for future speech-controlled wearables. Knowing that there have to be trade-offs, we want to illustrate the ideal requirements first and then derive a reasonable configuration from that. We want to emphasize that the following list is derived from our experience and feedback from users. Thus, it represents rather a basis for further discussion than a feature list of tomorrow's speech-controlled wearable computers. Table 3 summarizes the identified configuration.

Processing power: Speech engines still demand a minimum level of performance, which currently is equivalent to an Intel® Pentium® III, 500 MHz and 256 MB RAM [16][17].

Discharge cycle: Although working with hot-swappable batteries is acceptable in some applications, discharge cycles of at least one work shift with at least eight hours should be reached;

Size: Ideally, a user would not feel the wearable at all, i.e. it would be integrated in the clothing, such as smart textiles; given the previous requirements, it is more realistic to target a complete packaging that can be worn on the body (on a belt or in a pouch) and weighs less than 500 grams (which was felt as acceptable by users).

Connectivity: Based on the demands by the application, the system needs connectivity to peripherals and networking capabilities. Thus, standard connectors for cable-based peripherals, such as USB, and wireless connectivity, such as WLAN and Bluetooth, should be included – preferably as optional modules so as to reduce power consumption when not needed.

Display: Ideal would be unobtrusive displays that project the image in the visual field of the user or are integrated in clothing [15][18].

Table 3. Key features of proposed speech-controlled wearable

	Processing Power	**Battery Discharge Cycle**	**Physical Dimensions**	**Connectivity**	**Display**
Proposed speech-controlled wearable	Equivalent to Intel Pentium III 500 MHz or higher; minimum 256 MB RAM	minimum. 8hrs	Less than 500g including battery	USB; WLAN, Bluetooth	HMD or in clothing integrated display

4.4 Outlook: What is the Plan?

Xybernaut® will work on a speech-controlled wearable platform based on the identified requirements and the continuing feedback we get concerning hands-fee computing. In an effort to develop a "speech-ready" wearable, we will research, test and evaluate speech-related products (speech engines; software development kits; audio hardware, such as microphones, speakers, headsets; speech-enabled software). Based on a successful integration, we will recommend working system setups that are "speech-ready". For this process, Xybernaut® will seek cooperation with partners, such as speech technology providers; speech technology VARs; speech technology integrators; current and prospective customers; and universities and other research organizations.

5 Summary

Although probably the only hands-free computing platform currently available, speech-controlled wearables are still an emerging technology. This article showed some of the issue that have to be resolved and the research and development activities Xybernaut® performs as one of the major players in wearable computing. The goal for the "Speech-Controlled Wearable" has to be to cover many possible applications with one platform, but to limit the requirements in a way that a feasible and usable device can be configured. With starting a discussion on these requirements, we hope to get some fundamental input of ideas for the next generation of wearables that enable "Applied Wearable Computing".

References:

[1] Bürgy, Christian; *An Interaction Constraints Model for Mobile and Wearable Computer-Aided Engineering Systems in Industrial Applications*. Thesis, Dept. of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA, USA, 2002, p.14.

[2] Van Dam, Andries, *Post-WIMP User Interfaces*, Communications of the ACM, February 1997, Vol. 40, No. 2, pp.63-67.

[3] Bass, Len, *Interaction Technologies: Beyond the Desktop; Chapter in User Interfaces for All*, Lawrence Erlbaum Associaties, 2000, pp. 81-95.

[4] Smailagic, Asim, *An Evaluation of Audio-Centric CMU Wearable Computers*, ACM Journal on Special Topics in Mobile Networking and Applications, Vol. 6, No. 4, 1998, pp. 65-76.

[5] Meissner, Iris; *Development of a Concept for the Use of Mobile IT Devices for On-Site Data Collection Concerning Landfill Monitoring and the Prognosis of Landfill Reactions*; Thesis, Technical University, Darmstadt, Germany, 2001.

[6] http://www.vocollect.com retrieved 06. Jan. 2004.

[7] http://www.symbol.com/solutions/logistics/lekkerland.html retrieved 07. Jan. 2004.

[8] Smailagic, Asim., Siewiorek, Dan. P., *Application Design for Wearable and Context-Aware Computers*, IEEE Pervasive Computing, Vol. 1, No. 4, December 2002, pp. 20-29.

[9] http://www.symbol.com/products/barcode_scanners/barcode_wearable_wsc_ wss1000.html retrieved 12. Jan. 2004

[10] Piekarski, W., Smith, R., Wigley, G., Thomas, B., and Kearney, D., *Mobile Hand Tracking Using FPGAs for Low Powered Augmented Realit,*. Proceedings 8th International Symposium on Wearable Computers, Arlington, Va, Oct 2004.

[11] Deketelaere, Stephane, *How to take advantage of voice in wearable computer context*, Proceedings 1st International Forum on Applied Wearable Computing (IFAWC 2004), TZI-Bericht Nr. 30, 2004, pp. 85-96.

[12] Buergy, Christian, Garrett James H., *Using Business Process Modeling for the Integration of Wearable Computer Systems in MRO Applications*, Proceedings IEEE Mechatronics & Robotics 2004, pp. 1578-1583.

[13] Helbig Jörg, Schindler, Bernd, *Speech-Controlled Human-Machine Interaction*, IT- Information Technology (46) 6/2004, pp. 291-298.

[14] Bürgy, Christian; Garrett, James H., Jr; Klausner, Markus; Anlauf, Jürgen; Nobis, Günter; *Speech-controlled Wearable Computers for Automotive Shop Workers*; Proceedings to SAE World Congress, Detroit, MI, USA 2001.

[15] http://www.microvision.com/ retrieved 10. Jan. 2004.

[16] ftp://ftp.scansoft.com/pub/datasheets/DNS8_sdk.pdf retrieved 09. Jan. 2004.

[17] http://www.mediainterface.de retrieved 08. Jan. 2004.

[18] http://www.lumus-optical.com/ retrieved 10. Jan. 2004.