

A personal digital assistant as an advanced remote control for audio/video equipment

John de Vet & Vincent Buil

Philips Research
Prof. Holstlaan 4
5656 AA Eindhoven
The Netherlands
Email: {devet , builv }@natlab.research.philips.com

This paper describes a personal digital assistant that is used as a catalogue and advanced remote control to browse, select and play music in a compact disc jukebox. The application has been developed as a research prototype to identify advantages and disadvantages of different interaction styles for accessing large amounts of content. The basic concept provides easy access to a personal music catalogue, anywhere and anytime. It also allows you to control the CD jukebox. It employs a multimodal interaction style which combines voice control, touch input, visual output with animations and functional sounds. This helps to overcome the typical problem of accessing large information resources through small displays. In addition, redundancy in both input and output techniques offers people alternative ways of interacting with the content. The concept will be described and demonstrated, and relevant user studies will be explained.

Keywords: personal digital assistant, multimodal interaction style, voice control, compact disc jukebox, usability evaluation, personalisation

1. INTRODUCTION

A mobile personal device such as a personal digital assistant (PDA) provides good options to access large amounts of information and entertainment content anywhere and anytime. This paper describes a PDA that is used as a catalogue and advanced remote control to browse, select and play music tracks in a compact disc jukebox. The application has been developed as a research prototype to identify advantages and disadvantages of different interaction styles for accessing large amounts of content. It can also be used as a basis for identifying options for personalisation.

The basic concept employs a multimodal interaction style which combines voice control, touch input, visual output with animations, and functional sounds. The inclusion of both voice input and functional sounds helps to overcome the typical problem of accessing large information resources through small displays. Also, redundancy has been built in, in both input techniques as well as output techniques. This offers people alternative ways of interacting with the content, depending on context of use demands, on personal preferences, or on what is deemed socially appropriate. For example, selections

can be made by tapping an item in a list using the stylus, or by speaking the item's name directly. The last alternative would require a quiet environment, whereas the first alternative can be used in noisy environments.

The concept will be described and demonstrated, and relevant user studies will be explained

2. THE CONCEPT

A personal digital assistant is a handheld device that combines computing, communication, and networking features. It is typically pen-based, using a stylus rather than a keyboard for input, and offering handwriting recognition features. Some PDAs, such as the Philips Nino (Philips 1999), can also react to voice input by using voice recognition technologies.

The Philips Nino 300 has been used as a catalogue and remote control to select music compact discs in a personal CD jukebox. The CDs are shown in a list on the display of the PDA (see Figure 1). The list of CDs can be sorted by music style, artist name, release years and album names, by either using the stylus or voice commands. For example, the user can say

'Herbie Hancock', and the CDs of Herbie Hancock that are in the user's collection are shown on the PDA display. The first CD that is shown is highlighted. Simply saying 'play' results in activating the jukebox system to play the selected CD.

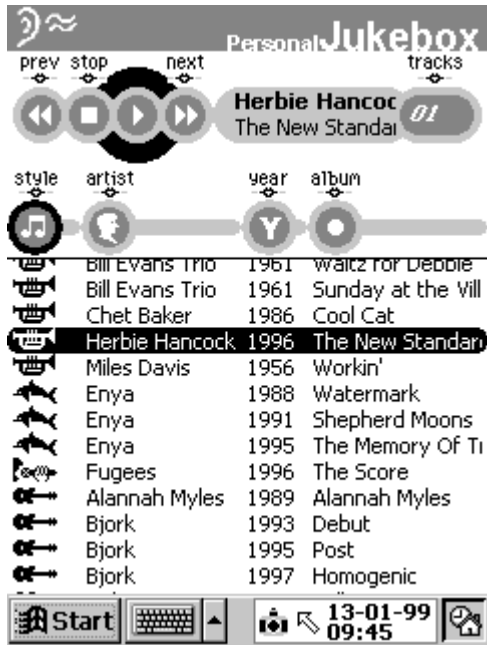


Figure 1: The PDA screen with the personal Jukebox user interface.

The information needed to create the CD catalogue on the PDA is simple: for each CD a number of attributes is available: artist, album, year, and style. This information can be downloaded from the Internet, for instance via CDDDB, a feature that most audio CD players on the PC now offer (CDDDB 1999). This means that the user does *not* need to enter this information manually, as is typically the case with current CD changers for the home. Ideally the jukebox system would send the ID information of the CDs to the PDA. Connecting the PDA to the PC would then result in an update of the catalogue. If the user has no connection to the Internet at home, it is still possible to enter the information (by typing on a PC keyboard, instead of pushing buttons on the changer).

The technology involved includes (see Figure 2):

- multimodal interaction (stylus gestures, voice input, animation, functional sound)
- Philips Vocon ASR software (continuous, word-based speaker dependent developed for small vocabulary and small 'footprint' (i.e. low memory & CPU resources) platforms, and hence cheaper devices.)
- infrared communication between PDA to PC via an IrDA (Infrared Data Association) link.

Figure 2: Set-up of PDA and PC simulation with IrDA transceiver on top of the left speaker.



- MP3 music on PC, meta-data from Internet. The CD changer is completely simulated on the PC using a modified Winamp MP3 player (Nullsoft 1999) and the CD collection is in MP3 format.

The following user benefits are anticipated:

- *Add-on remote control feature to an already bought product.*

A PDA is too expensive to be positioned as a personal remote control only, therefore the concept should be seen as an add-on feature. Existing universal remote controls, like are Marantz's RC2000 Mark II, Philips' Pronto and Sony's RM-AV2000, offer extensive and comparable control options. However, they do not offer the catalogue browsing option, which has been implemented on the PDA relatively easy.

- *Easy to use overviews of your CD collection on screen.*

The collection is shown on the display as a scrollable list of CD items, that can be sorted by music style, artist name, release year, or album title.

- *Using voice commands to access content directly.*

Music styles, artists, and release years can be named and immediately the associated subset of the collection will be shown on the display.

- *Browsing your CD catalogue anywhere and anytime.*

The catalogue can be shown to friends anywhere you are. Or you it can be consulted while shopping in your local CD store to see what you already have.

Anticipated user concerns are the following:

- *Getting the CD information on the PDA.*
This requires an Internet account to automatically download for instance CDDDB information (CDDDB 1999). The alternative would be for the user to manually type in the information. The catalogue in the current prototype is fixed and contained in a data file which can only be altered manually.
- *Adding a CD to your collection.*

Ideally, the catalogue could be updated when a new disc is inserted in the CD changer. Alternatively, the update of the catalogue would have to be done manually.

- *Training of voice commands.* Current word-based speech recognition technology requires training of new words, for example when a new CD is added. In the long term, phoneme-based, speaker independent speech recognition would be the solution, but this technology is not yet available on PDAs.

The opportunities that have been identified are:

- *Allows both personal use and group use.* A catalogue on a remote control can be used to find content of personal interest, without disturbing other people in the room who are using the audio/video equipment. The mobile device's display suits personal use. In case you want to enjoy audio or video together, i.e. for group use, a shared display (like a TV screen) would be better suited to find content of common interest.
- *Control multiple devices and a variety of content.*

The concept is also suitable for other applications, such as an electronic programme guide (EPG) that could be used as a personal TV programme recommender, or a catalogue of a videodisc (or videotape) collection. Hence it can offer access to a variety of content: music, TV programmes, film, theatre shows, sport events, and so on.

- *Hands-free control by voice.* For the control of audio/video equipment by voice, one controversial issue is the microphone location, and thus on how the automatic speech recognition (ASR) should take place. A microphone in the set (e.g. CD changer) allows for hands-free operation, but this scenario is more prone to noise interference, in particular to 'noise' coming from the audio/video equipment itself. A microphone in the remote control improves the quality of recognition, but does not free the hands. In case of a PDA, with on-board ASR and a reasonable display, the benefit of good visual feedback can compensate for the lack of hands-free operation. (When solely used for control, the PDA can be placed on the table, in principle, but the recognition will deteriorate.)

3. RESEARCH QUESTIONS

The research questions we had regarding *the concept* were:

- How do people appreciate the concept of using their organiser as a (universal) remote control for their audio/video equipment?
- How do people appreciate the concept of talking to a mobile device in the home or away?

Our research questions regarding *the user interface* were:

- Which operations are easier to perform with speech commands, and which operations are easier to perform on a touch screen?
- How to design a multimodal interaction style for use in different contexts (in the home, on the move, and away)?
- An organiser is a personal device, and thus can become a personal remote control that does not need to be shared with others. How can personalisation be exploited?

4. USER STUDIES

Our research group has conducted many user studies on the use of voice control in combination with other input techniques, for both stationary and portable products in the home environment. We have been most interested in relating user's conceptual operations to appropriate input and output techniques. Some of the findings will be summarised here.

4.1 Voice control

Operations that favour voice control:

- *Direct addressing of content:* Calling out names (e.g. of artists, categories, channels, etc.) is by far preferred over entering names with cursor keys on a remote control, or scrolling through names in a long list. Using voice commands is more natural and faster, and has better conceptual mapping (i.e. channel names vs. channel numbers). Earlier studies confirm that this is one of the main benefits of voice commands (e.g. the 'name dialling' feature in some mobile phones). However, for word-based speech recognition the names need to be trained in advance.
- *Menu navigation & selection:* Navigating through menu structures and selecting options is faster and preferably done with voice commands, compared to navigating with the cursor keys on a remote control. The task can be performed faster as there is no need to navigate stepwise through an option list or menu structure, and no need to switch attention back and forth between remote control and screen. Navigation through menu structures can be even more powerful with 'power commands', i.e. short cuts to options deeper in the menu structure, or macro functions that perform several selections at once (i.e. 'record this CD').
- *Setting a range:* When people have to set points on a scale, for example the start and stop time of a TV programme to be recorded on videotape, then voice commands are easier and faster to use than cursor keys. Setting times

with voice commands requires fewer actions than setting times on a slider bar with the cursor keys.

4.2 Manual control

Operations that favour manual control:

- *Scrolling in a long list:* Cursor keys are preferred and work faster for scrolling through long lists of content, if one does not know what one is looking for (browsing). Repeated voice commands like ‘up, up, up’ are annoying and slow, especially if the target item requires a lot of scrolling. An advantage of push buttons is that they can be held down for continuous scrolling. An even better alternative would be a real slider button or a rotary knob, as it facilitates controlling the position and displacement directly.

4.3 Voice and manual control combined

In one experiment we compared three versions of a Jukebox interface: voice input only, manual input only, and voice combined with manual input. We found that switching between voice and manual input seems unnatural to some users.

However, a combination of both input techniques can be very useful. For example, in the CD jukebox application on the PDA users can select a CD with the stylus, and subsequently invoke the ‘play’ command by voice.

Another advantage of combining voice and manual input, is that it provides alternative ways of operating the device. When automatic speech recognition is cumbersome, e.g. in a noisy environment or when the device is trained by another person, the manual input is a fallback option. User tests show that people want to have this possibility. Our post-experiment questionnaire results show that people really would want to use manual control instead of voice control in the following situations:

- *personal context:* when one is not in the mood to talk to a device, not able to talk (e.g. one has a hoarse voice), or when it is inappropriate (e.g. during a concert or presentation).
- *social context:* when one is talking to others, or when you don’t want to disturb other people in the room.
- *physical context:* when there is a lot of noise in the room – during a party for example – and voice control just doesn’t work very well.

5 DISCUSSION AND FUTURE RESEARCH

The concept presented in this paper is a prototype of what could be an add-on remote control feature for

people who already own a PDA. The concept is also suitable for other applications, such as an electronic programme guide (EPG) that could be used as a personal TV programme recommender, or a catalogue of your videotapes or videodiscs.

The disadvantages of mobile devices (small displays and few buttons, no keyboard) have been compensated by using voice input in combination with stylus input. Redundancy in the use of different input modalities makes it a robust interaction concept, that can be used in different contexts of use.

The real estate of the small display has been used in such a way that the items in the CD catalogue can be sorted on various attributes (artist, music style, release year), and sub-selections can be quickly made. In addition, animations and functional sounds have been added, to offer more redundancy in different output modalities.

This concept of a multimodal interaction style on a mobile device, seems also applicable to other domains than just entertainment, such as information and communication applications. It offers easy access to content through mobile devices. The mobile device does not necessarily store the content, although that would be possible, but it can be a gateway to that content, as exemplified by our application.

Our work has generated various questions for future research:

- *multi-user and multi-appliance:* A PDA is designed for personal use. How to design and implement voice control for use in a room with other people and other equipment?
- *shared interaction / scalability:* A single PDA does not support shared interaction: it is difficult to show your catalogue to others. A bigger screen that can be shared (e.g. a TV screen in the living room) is an option, but How well can a small-display application be scaled to a bigger displays?
- *personalisation:* Although the content, your CD collection, is personalised, the application and user interface are not. What are the options for personalising the personal remote control?

In the final paper, we will elaborate more on the experiments (design and data), on the advantages and disadvantages of the concept, and the implications for future research.

Acknowledgements

This work is a combined effort of our colleagues Vincent Buil, Berry Eggen, Luc Geurts, Paul Kaufholz, and Leon van Stuivenberg.

REFERENCES

CDDDB Online audio CD database (1999)
<http://www.cddb.com/>.

Philips Nino palm-PC official website (1999)
<http://nino.philips.com/>.
Nullsoft, Inc. Winamp music player (1999)
<http://www.winamp.com>