

# The Digital Dictaphone

## an exercise in audio-only interaction design

Lon Barfield (lon@design.nl)  
University of the West of England, General Design

### Overview

The paper that this extended abstract describes will discuss the current technological context for digital Dictaphones (a hand held, digital, sound storage device using complex and flexible sound structuring possibilities controlled using voice-recognition of a limited number of commands). It will then describe a project undertaken by over 200 students of Interaction Design to brainstorm and document design possibilities for such a system. Finally it will discuss the results and categorise them to build up a broad lexicon of ideas that can be further developed and investigated by the design community.

### The context

Until recently the miniaturisation of portable systems has been a struggle with miniaturisation of the internal technology. Who can forget the trend-setters at conferences straining with their first Apple Portables? Now however we are reaching the point where the technological barriers have been passed and we are up against a new and very different barrier; the interaction problems that miniaturisation brings. Keyboards below a certain size are unmanageable for large volumes of text input, new, more compact keyboards such as the one-hand chording approach of the MicroWriter/Agenda require considerable retraining [ref - Cykey keyboard], while small LCD screens struggle to display complex data in a space the size of a credit card.

One avenue that avoids these barriers is an audio/speech based interface. This has been a 'holy grail' system in the research world for many years, probably dating back to the portrayal of HAL in the film 2001. An ideal speech interface would require intelligence and error-free speech recognition, both of which are still beyond our reach. Even if these problems are solved there is still the problem of disturbance to others caused by audio interfaces (think of mobile phone users in public transport). However, earphones and advances in automated lip-reading could yet yield a silent voice interface.

There is a first, simpler step along this avenue; speech based interfaces that do not use AI and that do not have error free speech recognition. Imagine a Dictaphone where the sound is stored not on tape but in digital memory (these are already on the market and as memory prices drop their capacity rises). The big difference this will make is that the sound is not constrained to a linear tape but can be structured in new ways, (for example the user could record many small, separate 'sound-bites'). Couple this with simple voice-recognition of a limited number of commands and you have a 'hands-free' Dictaphone with complex and flexible sound structuring possibilities (a limited set of commands is necessitated because the system must learn the commands during a repetition/training phase). You could package such a system into a pen shape and wear it in the collar of your sweatshirt like Albert Einstein (something akin to Philips 'Magic Pen' in their Vision of the Future book [ref – Vision of the future, Philips]). Or you could put it around your wrist and use the 'finger in the ear' bone conduction method described at CHI'99 [ref – CHI 99 proceedings]. (Technically speaking a Dictaphone is an aid to dictation while what I am discussing here is an aid to recording and organising speech files)

The lack of buttons and visual display does tend to make the area somewhat intangible and less 'hot' than multicoloured data-visualisation. Appreciation of such systems depends to a great extent upon users immersing themselves in the interaction and actually using prototype systems. However there is a growing body of research being conducted in this area with the first experiments being carried out by Apple in their VoiceNotes project [ref - Interchi 93 proceedings]. Advances in this area will start to speed up as the Dictaphone market

merges with other related areas such as the new MP3 walkman systems, automated audio systems coupled to portable telephones and automated telephone information services.

## **The Digital Dictaphone at the Utrecht School of the Arts**

As well as being a practising interaction designer with a computer science background I also lecture on the Interaction Design course at the Utrecht School of the Arts (see below) [ref – Sigchi Bulletin july 97].

Designing the underlying model and interactions for such a digital Dictaphone is an exercise that I set the students to show them that interaction design is different from screen design. Since the machine is voice controlled and all feedback is audio feedback the students are constrained to think of the design of the underlying models and interaction and not to concern themselves with the external visual and tactile design of the product. No buttons, no lights, no screen, no strange three-dimensional shapes, just pure interaction design. It is a very unconventional design task but one that concentrates on the qualities that are needed for good user interface design. To emphasise the use of the designed system in a real-world context I asked the students to write a user-scenario; a short sketch or story involving someone using the system in a real-world situation. As well as making them think about how the system matches real-world user tasks it also made them concentrate on how rigorous their specification of the interactions is.

The reaction of the students to the project was interesting and varied. Almost all the design groups realised the problems with activating an interactive device using a voice command, how do you ensure that it is not activated accidentally? A similar problem was identified with other commands such as 'stop recording'.

The problem of designing the underlying structure of the sound bites and the commands that the user needed to interact with them was more complex and resulted in many different approaches. The most successful were those designs that gave each sound bite a short sound-title and had commands to quickly skip through the titles to help the user search for certain sound bites. However attaching titles leads to time overheads and conceptual overheads.

Another thing that became clear was that the user interface could be made easier to get to grips with if the command suite could be designed so that the same commands could be used in different contexts in the interaction.

Many students realised that it would be a simple matter to build a clock into the system and to give sound bites a time-stamp. This gives the user another axis to search along without involving them in the overheads of extra labelling activities. However, managing time in the input was difficult due to the high number of different commands necessary.

Time stamps could also be used for automatic 'house-keeping' systems to keep the collection of sound bites to a manageable size without requiring the user to undertake lengthy bouts of cleaning up, but the key question was how do you make the behaviour of automatic 'house-keeping' systems acceptable?

Other approaches included the problems of internationalising the system, the use of sounds and earcons and using error prone interaction techniques to speed up the interactions.

Setting the exercise for several years on the run to classes of 75 students resulted in a huge number of brainstorming hours being applied to the problems. This yielded many interesting approaches and identified further problems. It also produced systems that were undeniably bad, one example was the digital Dictaphone where the organisation of sound fragments was based on the metaphor of hard disk partitioning (something that of course all casual computer users are familiar with :-). Another student designed a suite of commands that didn't include a command to delete sound bites. Thus within a very short time the system would be filled to overflowing with sound files!

Solutions from this exercise are also highly applicable to the parallel market of automated telephone systems where users use an audio-based interface to access complex information structures. Is there anybody with a background in interaction design who has not got frustrated with endless repetitions of 'Which film would you like details about? Press 1 for 'What Dreams May Come', press 2 for 'The Never Ending Story, press 3 for

'Coma', press 4 for...'. Current Voice-Mail systems are also similar, they can be seen as digital Dictaphones where others create the sound bites and the user accesses and organises them.

Another important observation was that the speech nature of the interface makes the demoing of ideas during design very easy, designers can put a script together and then record it on a cassette recorder with two distinct voices. Compare this with the amount of work necessary to demo a visual based interface.

The written reports are a wealth of ideas and concepts produced by young minds at home with interactive technology and form a good starting point for brainstorming sessions involving small design teams.

## **Background**

The Interaction Design course at the Utrecht School of the Arts was one of the first four-year courses in the world dealing specifically with interaction design as an independent discipline. Now in it's 6th year, it's students are highly prized, and ex students are in employment at leading interaction companies such as BT, Apple, European Design Centre and Philips, indeed some national companies actually specify the course in their recruitment advertisements.

Lon Barfield is a research fellow at the University of the West of England and is associated with the internet solutions company General Design where he is a senior interaction designer involved in interactive web services. His lecturing activities include Interaction Design at the Utrecht School of the Arts and Industrial Design at the University of Delft. He is author of 'The User Interface, Concepts and Design' (Addison-Wesley) and has a regular usability column in the SIGCHI Bulletin.

(Dictaphone is a trade mark of Philips bv.)