

Talking Telephone Numbers

by
Dave Birch <<mailto:dave@chyp.com>>
Consult Hyperion <<http://www.chyp.com>>

Bill Gates, talking in Asia recently, made a point about the future of user interfaces and the limitations of keyboards and screens. One obvious interface is speech. The idea of talking to a computer rather than typing, pointing, pressing or touching it in any way is not only embedded in our shared cultural vision of the future (remember Hal?) but makes sense because it is so natural to us as a way of communicating. Following this line of thinking would lead one to suspect that voice-enabled e-commerce (ie, v-commerce) will become a common way to do business on line: the question is when?

It may not be so far away. The medium being the message, and all that, the greater part of this article was in fact dictated to my Macintosh laptop through IBM's splendid *ViaVoice* package. I estimate that the machine got the words right about 95% of the time: not perfect, by any means, but still a very convenient way to get some work done while both hands are otherwise occupied (looking through relevant documents, making a few notes and so on).

So voice recognition is starting to work, but how will it combine with the Net? There is already a standard for accessing web sites through the telephone (it's known as VoiceXML, or VXML) and there are already companies offering, or planning to offer, services in this area. The VoiceXML Forum, which was founded by AT&T, IBM, Lucent Technologies and Motorola recently completed Version 1.0 of the VoiceXML specification, which aims to make content more accessible via voice commands and 'traditional' phone interfaces [1].

Services like this mean that people will be able to follow up a web site and navigate their way through to find the information that they want using voice commands, and then have the information spoken back to them. However, navigating through vast amounts of information to find specific items probably works better through a web browser and large screen. Where the use of voice interfaces comes into its own is in this field of remote transactions. Phoning a bank to find an account balance or to initiate a bill payment is the perfect example.

At the heart of v-commerce are two rapidly evolving technologies: natural language speech recognition and speaker authentication. These will, in the medium term, make it possible for users to complete transactions all the way through from initial enquiries to giving payment details [2].

A voice recognition system would not have to be that comprehensive to satisfy these requirements. All it has to do is recognize that you are asking it to pay a bill and then present you with a list of pre-configured ways to pay (eg, "press 3 for Visa") to make it useful. Recognition technology is coming along nicely and some mobile operators have already implemented voice-recognition systems for customers to dial phone numbers or access voice-mail messages (eg, Orange's Wildfire) just by talking [3].

Another potentially important factor to consider is that to speech recognition may be combined with voice identification to add flexibility and security to interfaces. If

someone was able to phone their bank and say "give me my balance", it would be rather useful if the voice interface were to both recognise the customer's request and simultaneously authenticate the customer to the bank back-end systems. This obviates the need for passwords, and PIN numbers, or other additional authentication codes. Certainly it would be more convenient than navigating tiny mobile phone screens and punching data in using a tiny mobile phone keyboard.

The voice authentication systems that already in the market are capable of authenticating the user even if they have a cold or their voice is altered slightly. In the case of voice authentication through mobile handsets, the combination of possessing the handset and having your voice recognised should be more than sufficient for every day security purposes. If that happens, the new voice services could have a powerful effect on electronic commerce. Right now, 80% of people who begin a transaction on the Web cancel it before it's completed [4]. This can be because they can't be bothered to type in all of their details, because they can't understand the instructions or because it's just too much hassle: imagine just calling Amazon, saying what you want and then hanging up.

V-commerce isn't a niche. Once voice recognition and synthesis systems come in to use, their impact will be immediate and huge. Think of it this way: did Captain Kirk ever use a WAP interface on his communicator? Of course not: he spoke to it. That's what we're used to doing to phones, which is why voice-enabled services will actually be critical to the success of "wireless web" rather than just being a cute add-on [5].

Putting all of this together, then, indicates a medium term future for the digital mobile phone as a general purpose transaction device with the primary interface remaining voice. If, as some people suspect, wireless transactions could be responsible for almost half of all e-commerce in 3-5 years [6], then we'll all be talking telephone numbers.

Please note that an edited version of this article appeared in *The Guardian* (London) in the *Online* supplement (6th July 2000).

References

1. Moore, C. *VoiceXML specification to lead announcements at CT Expo* in *InfoWorld* (6th March 2000).
2. May, M. *V formation* in *Information Age*. **1**(37): p. 42-43 (January 2000).
3. Bainbridge, H. *New Talk* in *Wireless Review* (31st March 2000).
4. Rosenbush, S. *The Talking Internet* in *Business Week*: p. 174 (1st May 2000).
5. Kehoe, L. *Underdrawn by the wireless web* at *Financial Times* <www.ft.com> (on 29th March 2000).
6. Cole, A. *Surf your phone* in *The Industry Standard*: p. 84 (7th February 2000).