

TELEPHONE ACCESS FOR DEAF PEOPLE

Alistair D N Edwards

Department of Computer Science, University of York, U.K.

Abstract: The telephone is the most important piece of personal communication technology in the home. It is a technology that is accessible to nearly all people – except those with hearing impairments. Speech recognition and synthesis technology might be used to make telephone communication between a deaf and a hearing person possible. The necessary speaker-independent speech recognition technology is not currently available, but this paper reports a study in which such technology was simulated in order to test the feasibility of such communication. The results demonstrate that such a system would be highly desirable, but it will not be feasible until speech recognition rates are greatly improved.

Key words: Accessibility, deafness, telephone use.

1. INTRODUCTION

The telephone is probably the most used piece of every-day home communication technology. Unlike a lot of technology, it is accessible to almost all users – with the obvious exception of those with severe hearing impairments. The very usefulness and ubiquity of telephone use makes the handicapping effect of not having access that much greater. Improved technology could make telephones more accessible to people with hearing impairments, but there remains a question as to how useful such technology would be. This paper reports a small experiment in which automatic speech-to-text and text-to-speech technology was simulated in order to assess the utility of such a system.

2. THE PROBLEM

Of course there have been vast changes in telephone technology and usage in recent years. Greatest of these has been the introduction of mobile telephones. Their mobility is an important characteristic, but an additional feature that they introduced (almost as a side-effect – see Ocock, 2002) is that of text messaging (Short Messaging System, or SMS). This is probably the greatest innovation in telephony for deaf people. It adds a non-auditory – and thus accessible – channel to the telephone. Most importantly it is a feature available in *all* mobile phones¹³ as used by the vast majority of the population. In other words, every mobile phone user has the facility to ‘telephone’ deaf people – without having to acquire any additional equipment. This has been a boon to deaf people.

SMS messaging is not the same as telephone conversations, in that the communication is *asynchronous*. That is to say that it is not interactive. One person composes a message and sends it to their friend. The message is received on the recipient’s phone some time later (and there can be significant delays, because SMS is given a lower priority in the network than voice transmissions). The recipient will read the message at some time – but again there may be some delay. The recipient then decides whether to reply, and if so how (i.e. he or she may decide not to use SMS to reply, but to phone or go to see the person). This is very different from a telephone conversation, which takes place *synchronously*. That is to say that – as long as the recipient chooses to answer – then they are obliged to take part in a two-way conversation. This level of interactivity has a number of useful features: understanding and agreement can be negotiated quickly, for instance. (Reasons why people chose the asynchronous option of SMS over voice – and vice-versa – are discussed in Ocock, 2002).

Even though the use of mobile phones has affected the take up of landline telephones, the landline is likely to be around for some time yet. Thus, there still is an important role for voice telephony – for deaf and hearing people. For deaf users, the current solution to the inaccessibility of the voice telephone is a text-based alternative, based on the conventional telephone network. A *minicom*¹⁴ is a small terminal with a keyboard and a one-line text screen. Two minicom owners can connect their devices over the telephone network, so that the words typed by one can be read off the

¹³ This is not strictly true in that SMS was not a feature available in older generation systems, notably systems which were prevalent in the United States for some years.

¹⁴ Minicomms are also known as textphones, or – in the USA – as Telephone Devices for the Deaf or TDDs.

screen of the other. This enables two deaf people to communicate. (Figure 1).

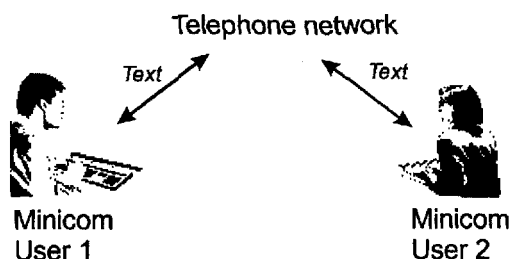


Figure 1. Two minicom users communicating. Both use the minicom keyboard and screen, connected through the conventional telephone network.

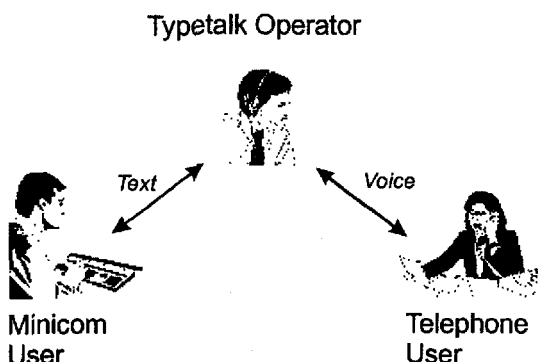


Figure 2. A Minicom User (deaf) communicating with a (hearing) Telephone User. This takes place via a human operator. He must type the spoken messages to the minicom user and read his input to the telephone user.

In principle the same system will work for a deaf person communicating with a hearing person – but few hearing people own minicomms. The solution in this case is to use a relay service. In such a system, two people (one with a minicom and the other with a phone) communicate via a intermediate human operator, as depicted in Figure 2. The words typed by the Minicom User are read out by the operator to the (hearing) Telephone User. She speaks her reply and that is typed back to the Minicom User by the operator. In the UK the *Typetalk* relay service is provided by the Royal National Institute for Deaf People (RNID)¹⁵.

¹⁵ <http://www.mid-typetalk.org.uk/>

Of course, there are a number of disadvantages to this solution, not the least that it is impossible to have a confidential conversation. It would be good, therefore, if the human intermediate operator could be replaced by technology. Speech recognition would be used to translate the spoken words into text, displayed on the minicom screen, and a speech synthesizer could pronounce the words that the deaf person typed to the hearing person. (Figure 3).

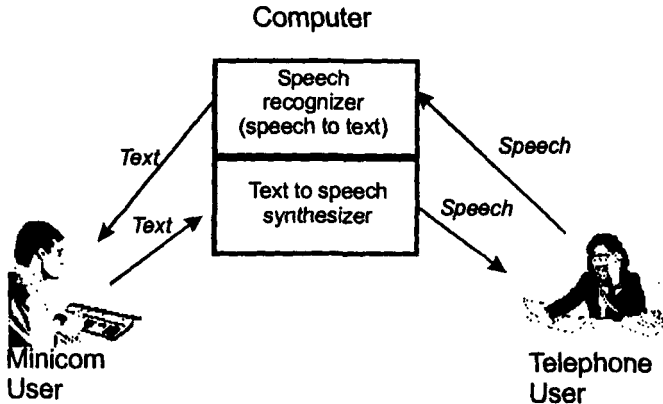


Figure 3. The ideal solution. The human operator is replaced by technology. A speech synthesizer will read out the text that the Minicom User has typed and a speech recognizer will convert the Telephone User's speech into text on the minicom screen.

While this solution has a lot of attractions, it is not currently feasible technically. In particular, the speech recognition engine for such a system would have to be capable of recognizing the speech of *any* person calling the system. In other words, the speech recognition must be *speaker-independent*. Current speech recognition software is quoted as having recognition rates of the order of 85-95% accuracy – but that applies only when the system has been trained to the individual's voice and manner of speaking. Training takes around 5-10 minutes, which would clearly be an impractical prelude to every phone call.

However, being optimistic, it must only be a matter of time before reliable speaker-independent speech recognition is achieved. The idea behind the study presented herein was to simulate that situation and thereby to test the likely effectiveness of the kind of facility depicted in Figure 3. It is often difficult to anticipate the effects – good and bad – of the introduction of new technology. Questions often remain unanswered, such as: Will users want (and pay for) the technology? How 'good' does the technology have to be in order to be successful? Will there be consequences of its adoption that are not anticipated? This is an example of a study which tries to avoid such

uncertainty, by simulating a technology which is not currently available – in this case speaker-independent speech recognition¹⁶.

3. SPEECH RECOGNITION

There have been great advances in recent years in speech recognition technology. Early systems had difficulty *segmenting* words, so that speakers had to insert an unnatural pause between each word spoken. Continuous speech recognition – which obviates the need for such pauses – has been a great advance. As mentioned above, recognition rates for such systems are quoted to be around 85-95%. At first glance this may seem to be technically impressive and sufficiently high to be useful. However, in practice such rates are too low to be useful in many applications; the time spent correcting errors and their effects outweighs the advantage of using speech recognition in the first place. On the other hand, it might be argued that in applications for which there is no alternative, such rates are acceptable. In other words, being able to recognize 85% of the words is better than 0%!

Another question is how and whether these rates are attainable. As will be shown below, in this experiment the rates were somewhat lower, much to the detriment of the results.

4. METHOD

The idea behind this experiment was to simulate the set-up depicted in Figure 3 and to carry out a number of conversations between a deaf minicom user and a hearing telephone user in order to find out whether the basic idea seemed viable.

The problem of speaker dependence was avoided by using a single person to act as the telephone user. In other words, before the experiment with both participants, the Telephone User had trained the (speaker-dependent) speech recognition package (*Dragon Point & Speak*, English Version 3.5¹⁷) to his voice. In principle, the speech output to the hearing Telephone User could have been generated using a speech synthesizer.

¹⁶ It might be said that this study continues the tradition of Dye *et al.* (1989). At a time when any form of automatic speech recognition was not available, they simulated such a system using a skilled human dictation-taker.

¹⁷ <http://www.scansoft.com/>

However, in practice it proved impossible to use a real speech synthesizer, and that had to be simulated by a person too (as explained below).

The system was simulated using two PC computers which were linked via a TCP/IP network and running *Lan Talk Pro* chat software¹⁸. The plan had been to use *Jaws*¹⁹ screen reading software²⁰ to read the text out to the Telephone User. However, there was a technical problem in that whenever a new line of text appeared on the screen, that line alone was not read, but rather the entire contents of the window. In other words, on every exchange the user would be forced to listen again to parts of the conversation they had already heard. It was therefore decided to simulate the speech synthesizer using a human reader. This had the added benefit that it effectively simulated a very high quality synthesizer. The computers were situated in separate rooms in the Department of Computer Science. The 'speech synthesizer' sat in the same room as the Telephone User who had no sight of the screen.

Both participants were given details of a number of role-play scenarios which they were to act out using the system:

- Scenario 1 – Arranging a meeting
- Scenario 2 – Purchasing cinema tickets
- Scenario 3 – Getting to know each other better

After the experiments, each of the participants was given a questionnaire to assess their reaction to the system.

5. PARTICIPANTS

The experiment was carried out twice. The Telephone User in both cases was the same. He had normal hearing and trained the recognition software to his speech. In Experiment 1 the Minicom User (referred to below as MU1) was an undergraduate student who was a Deaf sign-language user (British Sign Language, BSL) who had extensive experience of using minicomms (both in direct communication and through Typetalk). In Experiment 2 the Minicom User (MU2) was a professional sign language interpreter, who had been profoundly deaf, but now had a good level of hearing, thanks to the fitting of a cochlear implant (Tyler, 1993). Since she was born profoundly

¹⁸ <http://www.thaicybersoft.com/>

¹⁹ <http://www.freedomscientific.com/index.html/>

²⁰ A screen reader is a piece of software for blind users which presents text on a computer screen in synthetic speech. (Edwards, A. D. N. (1991). *Speech Synthesis: Technology for Disabled People*. London, Paul Chapman.)

deaf, she was brought up using BSL and taught to lip read. Despite having an implant, she still takes advantage of her lip-reading skill when conversing face-to-face with others. For telephone conversations she has to use a stereo speaker/amplifier attached to her telephone.

(1) Telephone User's speech

Hi, I've read your CV and I'm interested in interviewing you

I have read your CV and I would like to interview you

Can you hear me?

I'm free on Wednesday the 29th August for the entire day. Would that suit you?

Okay, does ten o'clock seem a reasonable time?

I think 10 in the morning would be good for me

Our offices are in the York Science Park

If you ask at the reception they should be able to direct you to where the interview will be taking place.

Okay, I shall see you on Wednesday 29th August at 10 o'clock in the Science Park.

Goodbye.

(2) Text displayed on Minicom screens

good afternoon, <Name>
it is <Name> here

*yesterday and then to higher eyes
ledger CV hands I mentioned
interviewing a further drop*

*i'm not sure if i understand you
woman's I have Rachel CV and I
would like to interview you*

KD here in

right, let me have a look at my diary
*I'm free on Wednesday the 29th
August for the entire day without
seeking*

*i'm also free on that day
okay does ten o'clock seem
reasonable time*

*what time would be appropriate for
you?*

*of 10 in the morning with a good for
main*

that's fine with me

*okay how offices are in the York
assigned sparkling.*

*if you Askew reception they should be
able to direct human to wear an
interviewer be taking place*

right, thank you

*okay I shall see you on Wednesday
29th August 10 o'clock in the science
park*

ok then, thanks. bye bye

Goodbye

Figure 4. Transcript of one of the conversations. Column 1 shows what the Telephone User said. Columns 2 shows how the conversation proceeded – as displayed on the screens – including misinterpretations. The Minicom User's (MU1) input – as spoken by the 'speech synthesizer' – is shown in plain typeface (and retains his original informal style of writing) while the translated Telephone User's speech is in italics.

6. RESULTS

The most significant result was that the speech recognition rate of the software was poor – despite the fact that it had been trained to the individual speaker. Recognition rates over the whole experiment, averaged 66%. Figure 4 shows the dialogue for one of the scenarios. It is apparent that there were many errors in the speech recognition. Some of them caused real misunderstanding, but for the most part the Minicom User was able to spot sufficient keywords to understand the intended meaning. It was noted that simple transcription errors were not the only problem. More subtle aspects of speech, such as prosody are also lost. Thus, for instance, 'Sorry?', intended as a request for clarification, appears as an apology when transcribed as 'Sorry'. Prosody is generally used as a means of signalling that the speaker has finished and is expecting a response. This was also lost, causing turn-taking errors and prompting the suggestion that there should be some kind of explicit signal that the speaker has finished the current utterance.

In the post-test questionnaires all participants agreed that the exchanges were time-consuming as a result of the frequent misinterpretations. However, according to the minicom users, the delays were insignificant in comparison to those experienced with Typetalk. For the Telephone User the delays were 'disconcerting' and 'frustrating'. He suggested that in a real phone call he would have discontinued the conversations in the belief that he was the victim of a crank call. His suggestion to eliminate such a problem was to notify the hearing person that a deaf person was on the line by using a 'canned' warning preceding the conversation.

MU1 was overwhelmingly positive about the system and appreciated that communication was directly between him and the Telephone User, that there was no third person involved in the conversation and total confidentiality could be assured. Again, in comparison to Typetalk, the system's speed was impressive – indeed, MU1 said, 'Speech recognition is like instant response. With Typetalk you wait for quite a while when the operator types the response'.

All participants agreed that they found the system to be a good idea and would use it in preference to Typetalk, subject to a number of conditions. All concurred that the delays meant the current system was appropriate for certain control situations, but quite unfeasible for conducting conversations and anything unstructured.

7. CONCLUSIONS AND RECOMMENDATIONS

It might be said that this experiment is flawed in that the component that turns out to have been the weakest link – the speech recognition – was precisely the part that was being simulated. However, the results do have important validity because they clearly highlight the fact that any (speaker-independent) speech recognition to be used in this application must be *very* reliable. That is to say that it will be unwise to attempt to introduce the technology at a point when it is merely adequate – and generate a negative backlash; better to wait until it is highly accurate and so acceptable. (Khine, 2001, lists specific recommendations as to the kind of improvements that would be most beneficial).

It might be argued that the experiment was too harsh, that too much was expected of the speech recognition. For instance, the results might have been better had more care been taken with the use of the microphone and training of the Telephone User. However, this would be artificial in that true conversations would be taking place using telephone equipment with untrained users.

The most important conclusion is that the proposed system does have great potential. The potential for truly confidential conversations cannot be under-estimated.

ACKNOWLEDGEMENTS

This paper is based on the work of Beatrice Khine, undertaken as the project component of her MSc (Khine, 2001). The invaluable assistance of the (anonymous) experiment participants is gratefully acknowledged.

REFERENCES

- Dye, R., Arnott, J. L., Newell, A. F., Carter, K. E. P. and Cruickshank, G. (1989). *Assessing the potential of future automatic speech recognition technology in text composition applications*. Proceedings of Simulation in the Development of User Interfaces, Brighton.
- Edwards, A. D. N. (1991). *Speech Synthesis: Technology for Disabled People*. London, Paul Chapman.
- Khine, B. H. (2001). *An Evaluation of Human Interaction Factors in Deaf Hearing Telephony*. MSc(IP) Project: University of York (<http://www.cs.york.ac.uk/%7Ealastair/projects/reports/pdf/khine.zip>).
- Ocock, M. (2002). *Why are text messages so popular?* York: University of York, Department of Computer Science Third-year Project Report.
- Tyler, R. S., (ed.) (1993). *Cochlear Implants: Audiological Foundations* San Diego: Singular.