AN EVALUATION OF HUMAN-INTERACTION FACTORS IN DEAF-HEARING TELEPHONY

BEATRICE H. H. KHINE
MSC. INFORMATION PROCESSING
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF YORK

SEPTEMBER 2001

SUPERVISOR: DR. ALISTAIR D. N. EDWARDS

	Report plus appendices	Report only
Word count:	20,012	14,514
Pages:	73	47

The counts were as counted by the Microsoft Word word count command.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to the following people:

- Dr. Alistair Edwards for his pastoral care, academic assistance and participation in the study;
- Ben Fletcher, Maggie Hogg and Richard Manley for their enthusiastic participation in the study;
- The Department of Computer Science's Support Group for setting up the system and network;
- My family and friends for their support and encouragement.

ABSTRACT

Deaf people currently communicate with hearing people using BT's Typetalk service where a human intermediary reads messages from the deaf person to the hearing person and types the hearing person's response to the deaf person. This raises questions about confidentiality and convenience.

Speech technology could provide a solution to these issues of confidentiality and convenience in the form of a system using a speech synthesiser to vocalise the deaf person's input and a speech recogniser to transcribe the hearing person's speech into text. Although this concept can be realised to a certain degree, the technology is still not developed enough to have true speaker independent speech recognition. However, this does not mean that issues related to human-interaction factors for such a system (when it is available) cannot be examined now.

This study sets out to uncover some of these factors through a simulation with a screen reader, speech recognition software and the necessary hardware. It will also involve deaf as well as hearing subjects who will run through a series of scenarios to interact with each other and the simulation system.

The objective is to make observations and draw conclusions, which will ultimately result in recommendations for improving the reliability of future speaker independent speech recognition in facilitating deaf-hearing telephony.

TABLE OF CONTENT

INTRODUCTION

For the deaf user the issue of telephony is contentious. At its simplest, it involves a text telephone to communicate with other deaf users. These telephones are known as Minicoms or Telephone Devices for the Deaf (TDD's) and are specially adapted telephones with small screens and keyboards. At its most convoluted, telephony is archaic, inconvenient and non-confidential for the deaf communicating with hearing persons through the Typetalk service, provided by BT. A proposed solution is to use speech technology – recognition and synthesis simultaneously – as a means of eliminating the inherent difficulties of direct communication between deaf and hearing persons. However, a system – in effect, an automatic Typetalk - employing the necessary technology is not available, and even if it were, there are still issues related to human interaction that would hinder optimal performance.

This study aims to provide some clues to the human-system and human-human interaction factors involved by simulating such a telephony system using a speaker dependent speech recogniser, chat software and speech synthesiser with the necessary hardware. The emphasis is not on delving into the technology behind the application, but on uncovering means to make such a speaker independent application truly reliable.

The study involved the use of subjects and, as mentioned above, the construction of a telephony system to simulate the appropriate conditions. Given the circumstances and time period of the study, the conditions were only approximations of real-life scenarios. Since the focus was on examining informal and semi-structured interactions, the specifications for each scenario were kept flexible.

The overall focus of the study was intended to be narrow, since it was limited to observing a telephony interaction between a profoundly deaf subject and a hearing subject. As explained later in this report, there are other categories of deafness, all of which have different characteristics and effects on communicational ability. The study was confined to just one classification of deafness for the sake of speed and scale. If the methodology was sound enough, then it could easily be 'reused' at a later date and applied to the other groups.

WHAT IS SPEECH?

Since "*natural speech can be seen as the ideal towards which speech synthesiser designers aspire*"[1] – that is, the benchmark against which all endeavours in speech technology are measured, an overview of speech - what, why and how? - are necessary. This introduction provides a brief outline of facts relating to human speech communication as a framework within which the principles behind speech technology is presented.

Speech represents the most natural means of communications among humans. It facilitates the immediate flow of complex ideas among interlocutors. In our infancy we learn to speak with little effort before we learn to read or write. So, it would not be overstating the case to assert that "*human communication is dominated by speech and hearing*"[2]. Its importance is underlined by the fact that "*it* [speech] *has been found in every human community that has been discovered*"[3]. In addition to speech, humans transfer data through writing (and reading), which allows for information to be given and received at a different time period or location. Despite these advantages writing is still "*less universal than speaking*"[4]. Also, speech captures certain nuances in meanings that may be inexpressible in writing where limitations are imposed by words and punctuation[5]. Nevertheless, speech is neither superior nor inferior to writing – both media have their own means of transferring ideas; they complement each other.

Generally, human speech generation begins with a coding process in the brain where ideas are translated into words. The brain then acts upon this process by formulating instructions for the vocal systems to follow and produce these words audibly. The employment of the vocal system involves mobilising a large number of organs to perform simultaneous actions.

Speech communication encompasses speech generation and recognition. In fact, there is always a transmitter and a receiver involved in all types of speech communication – either man or machine. The ideal speech recognition "machine" is a human listener.

---

1 Speech Synthesis: Technology for Disabled People, Edwards, A.D.N, page 12.
2 Speech Recognition by Machine, Ainsworth, W.A., page 2.
3 Speech Recognition by Machine, Ainsworth, W.A., page 2.
4 Speech Recognition by Machine, Ainsworth, W.A., page 2.
5 Speech Synthesis and Recognition, Holmes, J.N., page 2.

Speech is a "*modulated sequence of hisses and buzzes, the modulation being affected by the cavities and constrictions of the vocal tract*"[6]. It is characterised by the generation of individual sounds (also known as sound segments) or suprasegmental attributes. Segmental features convey "*the basic information of the utterance, the words and their message….the segmental features embody the sound of the words*"[7]. They can be segregated into two categories: voiced and unvoiced. Both sound types result from air continuously flowing through the vocal tract. The main difference between the two arises from whether this flow of air is subject to vibrations or not. In the case of voiced sounds, the air vibrates the vocal cords when passing through them as the speaker exhales - vowels are one example of voiced sounds. The speaker can control the pitch and type of sound. The former is managed by altering the tension in the vocal cords. While the latter is subject to the shape of cavities in the mouth through which the sound passes, for example, by moving the tongue and reshaping the lips[8]. Every individual has a natural pitch range or a normal vibration frequency, although this can be altered, albeit temporarily, through manipulating the tension in the vocal cords. An individual's gender is a determinant factor of her/his pitch – men in general have a lower pitch range than women and children. Unvoiced sounds rely on air passing through constrictions – *f* is an example of an unvoiced sound known as fricatives. Sounds may also be produced from blocking the airflow and, as above, can be categorised as being either voiced or unvoiced. Examples are *b, d, g* and *p, t, k* respectively.[9]

According to Dr. Edwards, suprasegmental features of speech refer to stress, rhythm, timing and intonation. He specifically discusses suprasegmental features in relation to prosody, which is the collective term for the aforementioned qualities and is "part of the information content of an utterance"[10]. Stress is to do with loudness and can be increased to accentuate a syllable. Rhythm and timing are linked as rhythm refers to the pattern of syllables while timing is the duration of a syllable (or phoneme) as well as of pauses within an utterance. And, intonation captures the pitch patterns (or 'tune' as Dr. Edwards calls them) of speech. An example of prosody in action is when a speaker converts a statement into a question by raising the pitch at the end of the statement.

[6] Tones of Voice: The Role of Intonation in Computer Speech Understanding, Longuet -Higgins, C., page 294.
[7] Speech Synthesis: Technology for Disabled People, Edwards, A.D.N., page 8.
[8] Speech Synthesis: Technology for Disabled People, Edwards, A.D.N., page 14.
[9] Speech Synthesis: Technology for Disabled People, Edwards, A.D.N., page 15.
[10] Speech Synthesis: Technology for Disabled People, Edwards, A.D.N., page 19.

In written text prosody is captured by punctuation, although Dr. Edwards highlights how ambiguous this can be - it is indicative rather than definitive and can leave the text open to variations in interpretations. For example, 'It's raining!' can said in anger or excitement – which is the correct version? The exclamation mark does not clarify the ambiguity[11]. This failure of punctuation impacts on speech recognition capabilities; in other words, "it limits the communication power of synthetic speech".[12] In addition to prosody, Dr. Edwards refers to another feature of language known as paralinguistics, which could have an effect on speech technology. Paralinguistic features give information about a speaker, such as the quality of voice indicating the speaker's emotional state[13], as well as including such things as whispering to indicate secrecy[14]. They are context-dependent, meaning that a set of global rules governing their production cannot be produced[15]. Finally, there are non-vocal speech qualifiers, as expressed in a speaker's body language. However, because this study relates to telephony and not face-to-face dialogue, body language will not count as a factor for consideration.

Bare details relating to what is an extremely complex function have only be given in order that this report is not overwhelmed by unnecessary, albeit highly interesting, linguistic material. The reader is presented with a very broad overview that should contain enough requisite facts to place and review this study in the correct context. Admittedly, it does not give a hint of the difficulties researchers have had in reproducing this organic phenomenon in silicon-based systems. But it is beyond the scope of this report to present in detail the fundamentals of speech, the history of speech technology or speculate on its future development.

WHY USE SPEECH FOR MAN-MACHINE COMMUNICATION?

Within the specific context of man-machine communication speech is "*potentially the fastest form of man-machine communication*".[16] W.A. Ainsworth puts the speaking rates for this type of communication as varying from 120 to 250 words per minute, which is faster than skilled typing rates. Its speed is not the only reason for speech being potentially the best communication medium between a machine and a man. It is acknowledged that speech is a highly appropriate for anything involving interactive dialogue or large volumes of data. Furthermore, he states, "*for all*

[11] Speech Synthesis: Technology for Disabled People, Edwards, A.D.N., page 18.
[12] Speech Synthesis: Technology for Disabled People, Edwards, A.D.N., page 19.
[13] Speech Synthesis: Technology for Disabled People, Edwards, A.D.N., page 19.
[14] Speech Synthesis and Recognition Systems, Yannakoudakis, E.J., Hutton, P.J., page 42.
[15] Speech Synthesis and Recognition Systems, Yannakoudakis, E.J., Hutton, P.J., page 42.
[16] Speech Recognition by Machine, Ainsworth, W.A., page 3.

applications where the machine is only accessible from a standard telephone instrument there is no practicable alternative".[17] The latter application is the one, which concerns this study.

Despite its advantages speech has not featured strongly in past man-machine communication – this distinction being held by typing – due to the machines' inability to understand speech[18]. As technology has advanced, speech has assumed a more prominent role in computer usage.

W.A. Ainsworth regards speech as having a number of distinct advantages over typing that allows it to be the primary channel of communication with systems, namely:

?? Faster problem solving

He reports of an experiment to determine which methods of communication were most effective in problem solving. The experiment involved two participants communicating by using a combination of various devices, e.g. microphone, typewriter. When speech was used, the time required to solve a problem was significantly shorter than with other methods. Furthermore, the number of messages the two participants needed to send to each other was fewer. The results were attributed to speech being the most natural form of communication as well as the participants being able to speak and think simultaneously.

?? Mobility

This advantage refers to the freedom speech gives to a system's user. The user retains full use of his hands and eyes to perform other tasks; an example is flying a plane where both hands and eyes are engaged. In this case, voice control would allow the pilot to give commands without disengaging his hands and eyes.

The second reference is to the '*omnidirectional*' quality of speech communication. That is, the speaker and listener are free to move and complete other tasks while still communicating by voice.

?? Minimal space

Speech eliminates the need for huge control panels on large and complex systems. Voice commands permit a system operator to control all functions from one position. The benefits

[17] Speech Recognition by Machine, Ainsworth, W.A., page 3.
[18] Speech Recognition by Machine, Ainsworth, W.A., page 2.

can also be seen in other locations and systems, such as in an aircraft cockpit, where voice control would release vital physical space.

?? Remote access

Speech is most effective in facilitating communication over long distances via the telephone network. A simple combination of telephone network and speech recognition grants access to systems not just at a local level, but also on a global scale.

An equally important note, which W.A. Ainsworth does not consider, but J.M Pierrel in his paper *Aspects of Man-Machine Voice Dialog* points out, is that speech makes use of natural language. The result is that users do not need to possess specialist knowledge in programming or some other form of formalised and learnt language in which to interact with their systems. Moreover, there is minimal need for sophisticated input-output hardware, such as microphone and loudspeaker respectively. Consequently, speech is a medium, which gives optimal access to systems for all users with the exception of those with certain impairments. Even so, this study is based on the premiss that sooner or later such conditions, including deafness, will not impede the application of speech as *modus operandi* for systems, telephonic or otherwise.

DEAFNESS

Since this is a study of deaf people and their means of communication, it would make comprehension of the results easier if the nebulous term deafness was defined and explained. The deafness in question is hearing impairment on a permanent basis. There will be no explanation or discussion of temporary hearing loss, which is also known as conductive hearing loss.

Contrary to the popular perception among hearing people there exist varying degrees of deafness. Deafness is a generic term that covers the whole spectrum of aural disabilities. It is used to represent all kinds of hearing loss that differ depending on the effects they have on the sufferers. Deafness ranges from mild to profound hearing loss and can affect people at any stage in life, or in other cases, be present from birth. The onset of deafness can be sudden or gradual.

Prior efforts have been expended in providing the definite definition of deafness, however, most have been vague. The best explanation asserts that someone is "*considered deaf if hearing impairment is so great, even with good amplification, that vision becomes the child's main link to the world and main channel of communication*"[19]. A more scientific definition measures the degree of hearing loss in decibels, ranging from mild to profound hearing loss. The same research that returned the aforementioned definition concluded that a loss of 35 to 75 dB was hard of hearing, while a loss of 90 dB or greater was exhibited by the profoundly deaf.

99% of all cases of permanent hearing impairment are sensorineural hearing loss, occurring exclusively in the inner ear and attributable to a number of causes. They include nerve damage arising from illness; meningitis; maternal rubella; RH blood incompatibility; scarlet fever; the absence of semicochlea fluid; cytomeglovirus; chronic exposure to loud noises (for heavy industry-related work); premature birth; head trauma; drug abuse and genetic disposition[20].

Despite the consensus on the generic definition, there fails to be agreement on the classifications of deafness. This disagreement owes itself to the very imprecise nature of the condition, as deafness in an individual "*lies in a continuum from a slight loss of hearing in one ear to total*

[19] Language & Deafness, Paul, Peter V., Quigley, Stephen P., 1984; quotes taken from http://home.inreach.com/torsi/typesofD.html

[20] *Types of Deafness, Deaf Education: A Parent's Guide* (anonymous) at http://homes.inreach.com/torsi/typesofD.html

deafness in both ears"[21]. At the same time, while one person may have mild hearing loss but experience extreme difficulty in communicating, another person may have a worse condition yet is a relatively proficient communicator. In other words, this contradiction highlights how subjective classifications can be. In an article entitled 'Deafness and Mental Health' John C. Denmark sought to dispel this vagary by producing a list of the conditions with their specific characteristics[22].

According to John C. Denmark there are those who are[23]:

?? Profoundly deaf

The profoundly deaf have little or no hearing. They are either born deaf or lose their hearing at an early stage of their childhood and grow up with sign language as their first language. Profound deafness can affect either one ear or both ears; in the former case a hearing aid may alleviate some of the hearing loss, but the latter case is beyond the benefit of a hearing aid.

?? Partial hearing

Those in this category experience mild, moderate, or severe hearing impairment prior to the acquisition of sufficient oral skills, resulting in their never having heard a word spoken correctly so impacting adversely on their oral communication skills. Hearing aids are of more benefit to this group than to the profoundly deaf, although like the profoundly deaf some people with partial hearing (depending on the severity of their condition) are adept at sign language.

?? Hard of hearing/ Deafened

This describes people who suffer either sudden or gradual hearing loss in adulthood. Known by some as hard of hearing and described by others as deafened, they have not grown up with sign language and will most likely have English as their mother tongue. Since the hearing loss occurs at a late stage of life, their speech patterns would be more developed than anyone from the previous two groups. This puts them at a disadvantage, as they are sometimes unable to

---

21 From a factsheet, *Deafness and Mental Health*, published by Mind: http://www.mind.org.uk/information/factsheets/D/deafness/Deafness_and_Mental_Health.asp

22 From a factsheet, *Deafness and Mental Health*, published by Mind: http://www.mind.org.uk/information/factsheets/D/deafness/Deafness_and_Mental_Health.asp

23 From a factsheet, *Deafness and Mental Health*, published by Mind: http://www.mind.org.uk/information/factsheets/D/deafness/Deafness_and_Mental_Health.asp

express themselves fluently in sign language. Those who are hard of hearing tend to identify themselves more closely with the hearing community than with other deaf people[24].

The consequences of having a hearing impairment on other faculties depend on the classification to which the sufferer belongs. John C. Denmark distinguishes between people with pre-lingual deafness, those with post-lingual or acquired deafness and people who are deaf-blind. The former group includes the profoundly deaf who have had no opportunity to acquire verbal language since they are unable to imitate sounds or understand verbal language from a small age (in some cases, from birth). This inability in verbal language also manifests itself in a lack of fluency in written language, thus written English becomes a second language and sign language is the mother tongue.

People with post-lingual deafness (also known as acquired deafness) suffer gradual or sudden deafness. Its effect on them is determined by the rate of its onset and its degree. Many, if not all, would have well-functioning verbal skills, which allow them to continue speaking, although their ability to monitor their speech for volume control and pitch etc. is impaired. A possible result is a significant deterioration in their speech, for which they can compensate by learning to lip-read.

The final group refers to people born with normal sight and hearing, who develop blindness and deafness. These conditions can arise simultaneously or at different times with the timing determining the residual skills that the sufferers retain. Those who become post-lingually deaf-blind still possess the ability to speak, and need only to use the deaf/blind manual alphabet or ‘block write’ on the palm of the hand to communicate. They also have the option of learning to read Braille. For those who are without literacy skills then reading by Braille or using the deaf-blind manual alphabet poses great difficulty and ultimately may hamper their communication with others.

People who are born blind and develop deafness at a later time have full spoken abilities and as with the post-linguistically deaf-blind can communicate using the deaf-blind alphabet, and read using Braille.

For the converse situation where people are profoundly deaf from birth and then become blind, sign language is the only feasible means of communication since the requisite literacy skills are

[24] From the Leeds Society for Blind and Deaf People website at http://www.hipleeds.org.uk/leeds.html

either underdeveloped or even undeveloped. Sign language allows the deaf-blind to receive information through a hands-on or visual frame interpreter. In effect, the deaf-blind 'shadows' "*the hands of the party that they are communicating with, or reduces the sign language to a reduced field of vision*"[25].

A mention of the languages the deaf community uses is necessary at this point, as they bear some relation to how fluid communication is between deaf and hearing persons. In the British Isles the deaf community (mainly pre-lingual) uses British Sign Language (BSL). It is comparable to English in that it has its own lexicon and syntax, but is distinct from spoken English. Another alternative, which borrows vocabulary from British Sign Language but shares grammar and structure with English, is Sign Supported English (SSE). The post-lingual deaf commonly uses it since it permits them to take advantage of the verbal and literacy skills that they already possess.

[25] From a factsheet, *Deafness and Mental Health*, published by Mind: http://www.mind.org.uk/information/factsheets/D/deafness/Deafness_and_Mental_Health.asp

CURRENT TELEPHONY SERVICES

Before venturing into conjectures and possibilities about synthetic speech telephony, it seems appropriate to provide a quick survey of existing means that the deaf have at their disposal for telephony.

Text telephony is considered the only commercially viable means for those who are deaf, hearing, impaired, deaf-blind or speech impaired to communicate over long distances. It takes advantage of the fact that most of the people lacking in hearing or speaking abilities are still able to write.

Text telephony enables two persons in different locations to communicate through the ordinary telephone network. All that is required is the use of specific hardware, namely two keyboards and screens, in order that text written by one person appears on the screen of the other person and vice versa. The interaction is real-time and instantaneous. It offers users, who are unable to use a voice telephone, the same telecommunication benefits that the general public enjoys in all circumstances, i.e. formal, private etc. Researchers emphasise how text telephony allows a "*feeling of contact and closeness to the person at the other end*" to develop in much the same way as voice telephony does for the hearing. This is achieved through "*immediate dialogue…*[making it possible] *to exchange fast comments in an easy way, to interrupt each other etc*."[26]

For a telephony system to be a text telephony system, it has to meet a number of basic criteria, identified by the Nordic Forum for Telecommunication and Disability. The Forum set out to draw up a communication protocol in response to what was felt to be a woeful lack of consideration given to human-system interaction issues of the disabled telephony user. They declared that the telephone network:

1. Should be able to support a direct connection between two parties for a written dialogue.

2. Should permit the transmission of both voice and text at the same time, especially 'voice over', which refers to the use of voice in one direction and written text in the other.

[26] Nordic Forum for Telecommunication and Disability (NFTH); data from http://www.stakes.fi/cost219/Texttelephony.html

3. Should support a relay service with a human operator. The service should allow users to talk to voice telephones and fulfill criterion 2 by using the operator as the contact with the text telephone user as well as the speaker with the hearing party. The service must also respect ethical rules regarding confidentiality, be available 24-hours per day while the waiting time to connect to the service should be as short as possible.

4. Should represent certain audio signals visually on the text telephone. Generally, all information that would be usually provided to the hearing user should be offered to the text telephone user as well.

5. Should provide add-on services, such as call-waiting, re-dial etc., as part of a comprehensive suite of functions.

Criterion 2 was intended to cater to the needs of those people who develop a hearing impairment at a late stage in life. For them the text telephone should be able to switch easily between voice and text since they would speak, but expect written text from the person at the other end of the line. The reverse is true for those with the contrary condition of being speech impaired but having fully functional hearing[27].

Despite its strength as a telecommunication medium for many deaf people, text telephony has a number of limitations, which prevent its mainstream use. In contrast to voice telephony, the two-way, interactive text communication is awkward and slow. According to the Instituut voor Doven (Netherlands), one of the many European institutions engaged in the UmptiDumpti project, it is also not suitable for confidential communication '*because one cannot identify the person on the other end by sight or voice*'[28]. Additionally there are the matters of economics and logistics, as adoption of such a service requires expensive specialist equipment on the part of the users – both hearing and deaf. Telecommunication companies providing text telephony have attempted to lessen the severity of the disadvantages by promoting text telephony as an extension of voice telephony. The resulting situation has ill-served the deaf community, as users are under the misconception that '*no special text telephones or standards are needed regarding text*

[27] The Nordic Forum for Telecommunication and Disability; data obtained from http://www.stakes.fi/cost219/Texttelephony.html

[28] UMPTIDUMPTI Project - Using Mobile Personal Telecommunications Innovation for the Disabled in UMTS Pervasive Integration; details were obtained from http://www.een.bris.ac.uk/UMPTIDUMPTI/data/deliverables/d21/Texttel3.html

telephony"[29]. Although this is not true and prevents investigations of the users' exact needs apart from the more obvious requirements.

In Britain, BT's Typetalk serves the deaf community. It is a national telephone relay service, established primarily for "*all deaf, deafblind, deafened, hard of hearing and speech-impaired people to communicate with hearing people anywhere in the world*"[30]. The service works by employing human operators to act as a link between the deaf user and the hearing person at the end of the line. Based on recent figures, more than 250,000 calls (of both business and personal natures) are made each month using Typetalk[31].

Typetalk uses a telephone with a built-in keyboard and screen – in other words, a textphone. Textphones relay conversations down a voice telephone line. If a call is to another textphone user, then the content appears on the screen of the textphone and Typetalk is not needed. It is when a deaf user wishes to contact a hearing person that Typetalk becomes involved. The Typetalk operator has two responsibilities: firstly, to relay what has been typed by the user. Secondly, to type in the other person's responses so that they appear on the deaf user's screen. The second procedure does not apply to those with speech impairments since they have the option of listening to the other person's response rather than having the operator type it in.

In theory Typetalk seems easy to use, but with an intermediary involved in the interaction, hearing persons dealing with Typetalk calls are reminded that they must:

- Speak carefully, clearly and slowly
- Spell out certain words to the operator
- Allow more time for each call
- Address the text user directly
- Use phrases "go ahead" or "over to you" after each section of speech
- Not interrupt
- Wait for the operator to give notice of when to speak again

From looking at the above list of instructions, it is immediately obvious whence Typetalk's limitations arise. All of the recommendations demonstrate how time-consuming conversations

[29] The Nordic Forum for Telecommunication and Disability; data obtained from http://www.stakes.fi/cost219/Texttelephony.html
[30] BT publicity brochures for Typetalk

via Typetalk can be as well as formulaic and controlled. Another issue is that of confidentiality. Informal voice telephone conversations are based on mutual trust and a certain level of intimacy between the interlocutors. Typetalk interaction is achieved through the presence of a third party that is a stranger. In recognition of this, BT highlights the service's comprehensive confidentiality and discreetness by making it clear that all operators are bound by the Official Secrets Act and Telecommunications Act. These Acts prevent the divulging of information by making such it a criminal offence.

On 4th July 2001 BT launched an improvement to its current Typetalk service, naming it TextDirect. Reacting to criticism about the cumbersome set-up procedure for calls using Typetalk, BT refined the relay service so that calls could be dialed directly to a text or voice telephone. No distinction would be made between the two types of telephony since the user inserts a five-digit prefix in front of any number whether it is for a text telephone or a voice telephone. Once contacted, the operator then would go on-line to participate in the rest of the call. Users are expected to benefit from the development on various fronts: the set-up process is quicker; there is no longer the need for an account number. Users receive combined bills for voice calls and Typetalk calls. And most importantly, textphone users are assured of "*virtually equal access*" to the telephone network.

This section has been an introduction to the most common type of telephony, which the deaf community has at its disposal to interact with hearing persons. There also exist mobile text telephony and video telephony[32], but since this study is focusing on fixed text telephony, these two alternative types of telephony are not discussed here.

[31] According to The Royal National Institute for Deaf People: http://www.rnid.org.uk/html/services_typetalk_home.html
[32] Please refer to Appendix D for information on current commercial developments in these two fields.

TECHNOLOGY

SPEECH RECOGNITION

This overview of speech recognition is comparatively longer than the following section on speech synthesis owing to the study's bias toward speech recognition. It is one application of speech recognition that will be the focus of the experiment; hence, it seems logical to pay more attention to general speech recognition as well as specific related details.

From birth humans are able to recognise speech and attain expert status at speech comprehension at a relatively young age – three years or so[33]. Speech recognition in computers is a wholly different matter, as what a computer does when it processes speech is different from what a human does. Ideally, most systems would recognise continuous words in a given language for a large vocabulary – a figure of 10,000 words is considered appropriate -, spoken by any person, regardless of accent (i.e. full speaker independence). In reality, current systems fall short of this ideal and are more likely to be speaker dependent or isolated word recognition systems since these are the easiest to be built. They are also commonly more adept at "*recognizing long words with many distinct features*"[34].

There are three different theoretical methodologies for achieving automatic speech recognition:

?? Acoustic-phonetic

The underlying principle is that all spoken words can be split into a finite group of phonetic units, which a system processes to determine which units have been spoken before decoding them into words.

?? Pattern recognition

This method is similar to the acoustic-phonetic approach in that it employs an algorithm to train a system to recognise patterns in specific words. However, it differs from the acoustic-phonetic in using a pattern recogniser with a neural net (or a hidden markov model) to search for patterns. There is no specific definition of patterns as phonetic units as in the acoustic-phonetic approach.

---

[33] Ability Hub Assistive Technology – Description of Speech Recognition: http://www.abilityhub.com/speech/speech-description.html

?? Artificial intelligence

This approach combines pattern recognition and acoustic-phonetic into a form of phonetic, syntactic, lexical and and/or semantic based analysis.

On a practical level the three theoretical approaches share a common starting point. They all initially rely on digital recordings being made of a speech sample. Thus, sound files recording the frequency and volume of the speech sample over time are created. Then, an analysis of the sound files is conducted in the following several stages:

1. Determine and label the state and end points of the utterances.
2. Filter the signal into frequency bands.
3. Segment the utterances into a fixed number.
4. Calculate the average of each band's data in each of the segments.

Only at the final step of the process does the system engage in any kind of sound recognition. Depending on the system's theoretical basis, the process is completed using the hidden markov model, a neural network or pattern-recognition. A system is trained on speech data, resulting in training sets of a certain number of repetitions per word being used for comparison with unknown words. In addition, the system can call upon extraneous linguistic and semantic data to establish the correctness of a word[35].

Compared to humans speech recognition in computers can be riddled with mistakes. Computers have a number of obstacles to surmount in listening to and understanding speech. The first is the correct separation and identification of speech from noise. Humans are able to filter noise easily and well, meaning that they can talk to each other almost anywhere and under any circumstances – for example, in busy train stations, across the dance floor, and in crowded restaurants[36].

Depending on the type of speech recognition technology in question, the second challenge involves 'training' the computer to distinguish between different speakers. Humans possess the

---

[34] Speech Recognition – Where are we, and where will we go? Mankoff, Jennifer, Georgia Tech College of Computing.
[35] Speech Recognition – Where are we, and where will we go? Mankoff, Jennifer, Georgia Tech College of Computing.
[36] Ability Hub Assistive Technology – Description of Speech Recognition: http://www.abilityhub.com/speech/speech-description.html

ability to discern between voices and do so all the time with great ease, so someone can hold simultaneous conversations with different people and know who is speaking. "Training" is an exercise that allows the speech recognition system to imitate this function. It involves the user "teaching" the system to recognise the user's voice prior to the first usage of the system, as well as adapting to the voice's pitch and volume.

The third consideration is how the systems should process two or more phrases that sound alike - an excellent example is "ice cream" and "I scream"[37]. A person would use the context and a dose of common sense to determine which phrase the speaker is actually saying. A computer system cannot rely on common sense since it does not truly understand what is being said, so it has to create a context by which to "navigate" through what is being said. It accomplishes this by tracking the frequency of words occurring by themselves and in conjunction with others. This information offers the computer several possibilities from which to choose the most likely word or phrase.

Finally, speech recognition systems have to overcome the natural tendency of humans to form incomplete sentences – it could be through mumbling, slurring or swallowing words. Humans assume that the listeners are capable of compensating for any of their oversights. But for systems mumbled speech or slurred words remain mumbled speech or slurred words since they only transcribe what is spoken, not what is eluded to, implied or assumed. As a result system users must speak evenly and clearly. Fortunately, accents should not pose a problem as long as the rules of speaking clearly and using standard grammar rules (for whichever language is applicable) are followed.

Types of speech recognition technology

The systems can be categorised into either speaker dependent, that is, they are able to handle large vocabularies for only one speaker or speaker independent, which deal with small vocabularies and a diverse group of speakers[38].

Speaker independent technology refers to systems "*that function regardless of the user base*". They are available in firms, which rely on providing direct and fast services to various consumers.

[37] Ability Hub Assistive Technology – Description of Speech Recognition: http://www.abilityhub.com/speech/speech-description.html

[38] Speech Recognition – Where are we, and where will we go? Mankoff, Jennifer, Georgia Tech College of Computing.

They require no "training" in recognising and adapting to the linguistic idiosyncrasies of each client. In fact, their strength is that they are able to process different speakers instantly[39].

By contrast, speaker dependent systems "*rely on the identity of the users*" and process large vocabularies for only one speaker[40]. They enable the user to achieve hands-free control and navigation of desktop applications simply through voice. They are designed for use by individuals therefore they require "training" with the user's voice and speech pattern before initial use. The training proceeds on a continuous basis with the program training itself over subsequent use. Over time and through frequent use the program familiarises itself with the user's speech and vocabulary (also known as speaker adaptive). Programs that recognise isolated words or continuous speech are also available.

The best examples of speaker dependent programs are the PC dictation software packages on the commercial market, such as Dragon System's NaturallySpeaking and IBM's ViaVoice.

Benefits of speech recognition

In general, regardless of the type of technology, the benefits of using speech recognition software are manifold. For the able-bodied user it "*increases productivity while decreasing the strain on the wrists and forearms*". In other words, it would assist in alleviating Repetitive Strain Injury. It also frees a user to undertake a host of tasks simultaneously, since all commands and anything that previously required hand movements can be completed by voice. A busy environment such as a hospital would be a perfect location for the said technology's application where users can work on computers whilst performing other tasks[41].

One purveyor of speech recognition software even claims that it "*can be useful for anyone who uses the keyboard on a regular basis. ….* [it] *lets you enter text faster than anyone can type*" – the assumption being that speech is a more natural and less self-conscious means of interacting with a system[42]. The user is liberated from the yoke of typing, which has over the years become a necessary skill when dealing with systems. Furthermore, when speech recognition is combined

---

39 Technology evaluation, completed by Carla D. Pinon at Stanford University: http://www.stanford.edu/~cpinon/speech.html
40 Speech Recognition – Where are we, and where will we go? Mankoff, Jennifer, Georgia Tech College of Computing.
41 Speech Recognition – Where are we, and where will we go? Mankoff, Jennifer, Georgia Tech College of Computing.
42 Who should use speech recognition technology? 1stVoice, resellers of Dragon software, from: http://www.1stvoice.com/users.html

with natural language understanding, users no longer need to be aware of the technical details involved in using systems to have access to them[43].

Speech recognition has already been marked as a technology that has, when applied appropriately, is of immense value to those with disabilities. As such, it is known as an assistive technology and considered a natural and easy method of accessing the computer, especially for those who cannot move a mouse or type on a keyboard, such as paraplegics.

Problems with speech recognition

The disadvantages lie in the very thing that defines speech recognition: speech. Obviously, those who are unable to speak or cannot enunciate well are at a disadvantage with this technology. As mentioned above, speech recognition systems 'mis-process' slurred words so any use they have for disabled users is undermined, since those with speech impairments are particularly affected.

Users have to display a certain degree of mental dexterity in co-ordinating their diction and other commands. Users must learn to give verbal commands to format, edit and correct any text. If the user is unaccustomed to dictation, then the technology becomes ineffective in completing complex tasks. An example is the use of a spreadsheet program to build a complicated financial model where cells are referenced by multiple ranges. It is easier to perform the task using the keyboard or mouse rather than a set of spoken instructions. The same applies to word processing programs when correcting and editing documents can become frustrating in comparison to creating the documents. The reason for this disparity in performance is that humans speak faster than they can type, but they navigate and use their hands faster than they can issue instructions44.

The aforementioned technical constraints have minimised the impact of speech recognition technology on mainstream means of communication. However, steps are being taken to remedy the situation. One such move has been to develop 'intelligent language models' to bridge the gap in performances between spoken instructions and hand navigation in completing tasks and improve usability. The aim is to make programs understand natural commands, such as "bold the last three words"45, instead of navigating to the bcation of the three words then issuing an instruction to highlight them before bolding them. In addition, the 'intelligent language models'

[43] Speech Recognition – Where are we, and where will we go? Mankoff, Jennifer, Georgia Tech College of Computing.

[44] Technology evaluation, completed by Carla D. Pinon at Stanford University: http://www.stanford.edu/~cpinon/speech.html

[45] Technology evaluation, completed by Carla D. Pinon at Stanford University: http://www.stanford.edu/~cpinon/speech.html

serve to increase the accuracy rates in order that, for example, a 1,000-word document only contains the minimal number of error words. It should also allow for the automatisation of capitalisation and punctuation to assist in any correcting or editing exercise.

Applications of speech recognition

Depending on whether they are speaker dependent or independent, current speech recognition systems are used for:

- Dictation- translation of the spoken word into written text.
  This application makes speech recognition a form of input, appropriate for dictation software and menu selection. There are already telephones that use speech input instead of number-pad menus[46]. Word processing is the forte of dictation programs and many dictation programs exist for generic use as well as specialised application in professional fields, such as medicine or law.

- Computer Control- control of the computer, and software applications by speaking commands[47].
  A commercial example is an automated information service, employed by an American airline company to provide callers with flight information[48].

- Parsing tool – easier searching and indexing of recorded audio and video data[49].

SPEECH SYNTHESIS

Speech synthesis is the output of speech from a computer system, although it is not just concerned with the generation of sounds, but also with the control of their production. The systems can use mechanical and electronic synthesis to produce sound, but for the purpose of this study the focus will be on electronic synthesis methods. Dr. Edwards in his book 'Speech Synthesis: Technology for Disabled People' identifies two categories: copy synthesis and synthesis-by-rule.

---

[46] Speech Recognition – Where are we, and where will we go? Mankoff, Jennifer, Georgia Tech College of Computing.

[47] Ability Hub Assistive Technology – Description of Speech Recognition: http://www.abilityhub.com/speech/speech-description.html

[48] Technology evaluation, completed by Carla D. Pinon at Stanford University: http://www.stanford.edu/~cpinon/speech.html

[49] Speech Recognition – Where are we, and where will we go? Mankoff, Jennifer, Georgia Tech College of Computing.

John P. Cater provides a good analogy to explain the difference between the two by comparing copy synthesis to a 'photograph' of human speech; while he considers synthesis-by-rule as an 'oil painting' of the same speech. He states that a photograph is a reproduction of the original and an oil painting is very close to the original, but will never be the same[50].

Copy synthesis involves the reconstruction of direct speech; it is comparable to a digital recorder recording the output from a human speaker, saving it to memory and then processing or reconstructing it as and when it is needed. The system relies on pre-spoken and pre-recorded vocabulary, which can comprise of whole sentences, phrases or words or even phonemes[51]. Once these are encoded and stored, the system is at liberty to simply string them together in a manner that attempts to approximate normal human articulation. The 'digitalised recording' is done through the use of a sampling mechanism that continually measures the amplitude of a speech waveform at specific intervals. Upon completion of the sampling, a digital amplitude record of the originally spoken sound then exists in memory to be recalled at random in a manner analogous to "*written words of a text…rapidly accessed by a word processor*"[52].

It is with the sampling mechanism that copy synthesis fragments into different approaches. Firstly, there is the direct digitalisation of sound, characterised by a high sampling rate and high memory requirement, and known as pulse code modulation (PCM) (or waveform encoding or direct speech reconstruction technique).

Secondly, delta modulation (also known as adaptive delta pulse code modulation) is similar to PCM, its one difference being its measurement of relative amplitude changes rather than absolute amplitude values. It is an incremental encoding method[53] because it stores "*the difference between successive samples than their absolute values*"[54]. As a result, the number of storage bits for each sample is reduced, thus delta modulation is a more efficient method of digital encoding than PCM in terms of memory requirements.

The final method of using pre-recorded sounds for computer speech output is linear predictive coding (LPC). As the term suggests, it is based on the predictability of speech waveforms and is

---

50 Electronically Speaking: Computer Speech Generation by John P. Cater, page 101.
51 Speech Synthesis: Technology for Disabled People, Edwards, A.D.N, page 24.
52 Speech Synthesis: Technology for Disabled People, Edwards, A.D.N, page 23.
53 Electronically Speaking: Computer Speech Generation by John P. Cater, page 96.
54 Speech Synthesis: Technology for Disabled People, Edwards, A.D.N, page 24.

more complex to understand than the previous two approaches. Despite it being based on a speaker speaking words and to a certain extent incorporating the pulse coded modulation process (which converts the human speech signal into digital samples), it is mathematical in nature. Dr. Edwards explains it succinctly when he wrote that *"it is possible to show that the value of (say) the hundredth sample (S_{100}) can be predicted from the previous fifty samples (S_{50}….S_{99}), so that it is not necessary to store the hundredth sample (S_{100}). In other words, the information contained in S_{100} is redundant"*[55]. Such redundancy has the positive effect of making LPC an even better method of data compression than delta modulation.

When speech waveform is processed and reproduced in analog format, it is subject to distortion and noise. However, digitalisation discounts these effects and at the same time allows speech waveforms transferred from device to device to progress in an unchanged form. That is, digitalisation makes exact copying of the sound by each device (for example, a microphone) possible[56]. The compound result is that synthesisers constructed along the copy synthesis principle produce high quality speech that sounds human.

Synthesis-by-rule is 'real' speech synthesis owing to the absence of pre-recorded human speech. It is driven by the analog formant frequency method, based on phonemes and the phonetic breakdown of a particular language. Essentially, the method uses an electronic circuit – an analog formant synthesiser - with as many bandpass filters as necessary to filter out a host of electrically generated formant frequencies (representing different phoneme sounds). An electrical speaker then converts the filtered signal into audible speech. This synthesiser can be adapted and extended to produce vowels, fricatives and stops as well as nasal consonants by manipulating the filters, which act as controls. The manipulation can be manual or, with relevance to this study, by computer through a specially written program. The program breaks an input word, such as 'hello', into individual component phonemes (parse) of 'h', 'eh', 'l', and 'o', before generating the individual phonemes and putting them together to form the complete word and pushing it out through a speaker[57].

Problems with speech synthesis

The degree of understandability and the quality of speech production are factors in assessing a speech synthesizer's performance. So, no one synthesisation principle is better than the other at

---

[55] Speech Synthesis: Technology for Disabled People, Edwards, A.D.N, page 24.
[56] Speech Synthesis: Technology for Disabled People, Edwards, A.D.N, page 21.

producing synthesisers with excellent performance since it seems to depend on which factor is being considered at the time.

Copy synthesis synthesizers possess limited vocabulary, since they are constrained by the number of recorded units of human speech in the memory – words cannot be constructed if they have not been pre-recorded!

For synthesis-by-rule synthesizers, the converse is true. Theoretically, the recordings of phonemes should enable them to boast of an unlimited vocabulary, however, it comes at a price, namely a poorer quality of speech.

Applications of speech synthesis

Dr. Edwards identifies two groups of users who benefit from speech synthesis technology. The first group is those "*people who lack the physical ability to produce speech. For them, synthetic speech can become their voice via a speech-based communicator*"[58]. He refers to them as 'expressive' synthetic speech users. In context of this study, the deaf subjects would fall into this category since the system set-up demands that synthetic speech is used to achieve telephony with their hearing counterparts, regardless of their own natural ability to produce speech. Conversely, the hearing counterparts belong to the second group who would "*use it* [speech synthesis] *as a means of receiving information, which might otherwise be presented in a written form. In particular it is a means of access to computer output*"[59]. They are "receptive users" as speech synthesis represents a "*means of receiving information from computers*" – in this case, from other users.

Apart from its role as an essential aid to those with speech and visual impairments, speech synthesis has additional uses; "*some are little more than gimmicks, such as the car which talks to the driver, saying when his or her seat belt is not fastened, for instance*"[60]. Others, however, include telephone response systems that are capable of automatically quoting stock prices and bank balances[61].

---

[57] Electronically Speaking: Computer Speech Generation by John P. Cater, page 109.
[58] Speech Synthesis: Technology for Disabled People, Edwards, A.D.N, page 1.
[59] Speech Synthesis: Technology for Disabled People, Edwards, A.D.N, page 2.
[60] Speech Synthesis: Technology for Disabled People, Edwards, A.D.N, page 1.
[61] Electronically Speaking: Computer Speech Generation by John P. Cater, page 19.

METHODOLOGY

This section underwent a number of changes before the final methodology was adopted. The basic concept was to enable real-time telephony through speech recognition and synthesis technology. Apart from the speech technology software, the conceptual model would involve two computers representing the two ends of a physical telephonic connection and messaging software to provide the virtual link between the computers. At one end there would be the deaf subject, at the other would be the hearing subject in a different location. The messaging software would ensure that the telephony was real-time and instantaneous, while the hearing subject would 'listen' to and 'talk' with the deaf party through the aid of the speech synthesiser and speech recognition software respectively.

The original conceptual model was a generic and high-level overview of how the system should function. There were no strict technical specifications, meaning that the next stage in the development of the final methodology involved searching for appropriate hardware and compatible software, which would determine how the final system would actually work.

Owing to the limitations imposed by time and resources, the experiment was to be carried out on campus. The chosen site was the Human Computer Interaction Laboratory and the adjacent room – the deaf subject and his/her terminal were to be in the HCI lab and the hearing subject was to be located in the adjacent room. Each subject was to have a keyboard and a monitor, however the hearing subject was not expected to utilise his terminal, as this would have reduced the experiment to being a one-to-one computer chat via keyboard and screen, rather than an attempt at telephony.

A screen reader program was to be installed in the memory of the hearing subject's computer along with speech recognition software. Thus, he would 'hear' (through his headset) what his deaf counterpart was typing out at the other end – text to speech conversion – and he would respond using the speech recogniser (via his microphone headset) – speech to text conversion. Whatever was transcribed on the monitor of the hearing subject would appear on the deaf subject's screen; this facility was courtesy of the messaging software resident in both computers.

Once this system was in place, the subjects would begin undertaking a series of semi-structured exercises. The purpose was not to evaluate the performance of the system, but to investigate any human interaction issues related to this form of telephony through the scenarios. Each subject was to be given an outline of each role s/he was expected to play, and the context in which the telephone conversation was to take place. The subjects were to be given leeway in determining the exact details and direction of each conversation; the only overriding criterion being that the exercise was semi-structured.

In total there were to be three role-play scenarios: job interview, purchasing a cinema ticket and a personal chat. With the exception of the final scenario, each scenario had two versions – version one was to be the straightforward run of events; while version two would involve the resolution of a problem, e.g. one of the subjects would have a clash in her/his diary for the job interview. This was to prompt a discussion with the other party to find a better date and time. What was of paramount concern was not the length or the nature of the conversations, but their approximation to real-life situations and their 'naturalness'.

CONSTRAINTS OF THE STUDY

Initially the study had difficulties in finding deaf subjects (from any of the three categories of deafness) willing to participate in the role-plays. A letter with full details of the study was sent to the York Deaf Society to request for volunteers. No response was forthcoming. One of the possible reasons for this lack of co-operation could be attributable to the deaf community's desire to maintain its independence and its perceived unwillingness to integrate into mainstream (and hearing) society. The study with its emphasis on facilitating direct communication between the two communities would have gone some way to compromising this independence and represented another step on the path toward integration. It became clear that establishing meaningful contacts with the deaf community would take longer than the gestation period of this report.

More likely, however, was that the tasks involved in the study were difficult and complex, especially in light of certain general comments made by session two's Subject 1. According to her, the tasks were beyond the comprehension of the average profoundly deaf person. Not just in terms in of the technical details specified, but also the other skills required – familiarity of situations approximating the role-play scenarios, a level of education sufficient to understanding the study's key concepts, a developed vocabulary, the social confidence to deal with 'a stranger' on the other end etc.

Many assumptions were made regarding the technical knowledge and level of English for the deaf subjects, which may have affected the validity of the data gathered. Indeed, session one's Subject 1 can be considered an exception among the profoundly deaf, owing as much to his naturally high intellect as to his family background. Unlike many profoundly deaf people, Subject 1's immediate family members – father, mother and sister – all learnt sign language. Usually there is only one family member who communicates with the profoundly deaf person. In the case of Subject 1, his education was as inclusive as possible with input from his whole family. He is a life-long BSL user as he was encouraged to sign rather than lip-read throughout his childhood. He was not isolated from the general hearing community and as a result he possesses social skills and experiences equitable with those of any hearing person. His reading age surpasses the national average reading age for the profoundly deaf, which is eight years old; it is in fact comparable to that of any hearing university undergraduate. His vocabulary is extensive and definitely contributed to his ability to compensate for some of the misunderstandings that arose as a result of the system's mis-transcriptions. It enabled him to improvise easily in his role-plays and converse meaningfully with Subject 2, which could have been too much for another profoundly deaf candidate who may have had limited life experience and consequently have been too uncomfortable with the idea of interacting with a 'complete' stranger. With Subject 1 (session one) there was no hesitation in reading and he demonstrated no difficulty with any of the terms that appeared during the conversations. His general technical knowledge and familiarity with the technology employed in the study also meant that he did not require training or in-depth explanations. If another profoundly deaf subject had been used, more time and possibly a BSL interpreter may have been required to provide more guidance and run through the whole procedure and set-up in terms with which that subject could feel comfortable and comprehend. Supposedly, the concept of speech recognition would be beyond the understanding of the average profoundly deaf person since s/he would have no experience of using her/his voice.

The choice of software was also affected, since the original speech synthesiser to be used was JAWS but owing to its inflexibility Dr. Edwards had to assume the role of screen reader. There was no time to test out other screen readers, which might have been able to read out new text (in the dialog box) rather than the whole conversation from beginning to end as JAWS did, as well as be compatible with scrollable dialog boxes. The speech recogniser was not exempt either from this limitation on resources. Dragon Point & Speak, despite being finally selected, was problematic and proved erratic during its training, rejecting all attempts to train it to a female voice. It could possible have been a consequence of Dragon Point & Speaker's optimisation for American accents. In the end the use of a male speaker for Subject 2 was justified, but it would still have been useful to have a range of options.

Another major constraint was the lack of time. It had limited the search for a 'typical' profoundly deaf person, but it also meant the study could not be expanded to encompass a more diverse group of hearing and deaf subjects. The most obvious impact was on the conclusions to be drawn from any result. In other words, generalisations applying to deaf-hearing communication could not be made since the subject sample was biased. Both Subject 1s were exceptions and accordingly, any outcome would be a reflection of their exceptionality rather than of the average profoundly deaf person.

SPECIFICATIONS

A breakdown of the study's specifications as well as the details of the scenarios to be enacted are given as follows, in order that this study can be replicated for each of the classifications of deafness.

Screen reader

Generally, a screen reader is a software program that makes standard software accessible to blind users by speaking out the content of the active screen – hence, the term 'screen reader' – through a speech synthesiser. It offers blind users greater user equality and access to standard, visual software in a number of ways. Firstly, its voice output technology 'reads' out text on the screen. At the same time, it is able to identify some graphics on the screen, such as the toolbar buttons. Secondly, it allows the user to have control over the computer using the keyboard rather than the mouse[62]. The keyboard commands are the same as those used by the sighted plus special combinations of keys provided by the screen reader.[63]. Only one screen reader is needed since "*one screen reader give access to (nearly) all applications*"[64].

Originally a demo version of JAWS for Windows 3.7 (as developed by Henter-Joyce) was selected as the screen reader. However, it failed to perform exactly to the system's requirements

---

62 How does a screen reader work? From the Iowa Department for the Blind website: http://www.blind.state.ia.us/access/how.html
63 How does a screen reader work? From the Iowa Department for the Blind website: http://www.blind.state.ia.us/access/how.html

during a test run. Whereas the idea was to have JAWS 'speak' every new text response that appeared in the dialog box throughout the sessions, the reality was that JAWS repeated each scenario from the beginning to the end, making the interaction too artificial for the purpose of the study. A substitute for JAWS was considered – WillowTalk by Willow Pond Software – but this program too was found to be inadequate for the study's needs. In the end, Dr. Edwards, the supervisor for the study, simulated the functions of a future, high-level speech synthesiser by reading out Subject 1's responses.

Speech recognition program

Dragon Point & Speak was selected to provide the system's speech recognition facility. Dragon Systems claims that it allows the user direct dictation into '*virtually any Windows application that accepts text*' – that is, it can be used with word processors or other programs. Point & Speak recognizes natural and continuous speech.

The user needs only to activate any window or text box (by clicking in it) before dictating. Point & Speak converts what has been said into text in the active window or text box. It also allows different editing means so that a user can correct and modify any work that has been done with the mouse and keyboard or through voice commands. Furthermore, the voice commands enable control of menus and dialog box options

Chatware

In order to establish a virtual connection for text data transfer between the two terminals in real time, a viable piece of chat software was needed. A total of six shareware programs for instant messaging was downloaded and tested for compatibility with the system set up as well as usability. Most were eliminated as they could not operate via a local network or needed an open Internet connection. To be considered for inclusion in the study, the applications had to be appropriate for a stand-alone network (basically LAN-based), possess chat logging capability and be able to run on Windows 98. Whereas the other shareware were problematic when installed, *Cool Chat* and *Lan Talk Pro* satisfied all criteria and were relatively easy to download and install.

According to its publisher, NetAbyss, *Cool Chat* allows users to chat via the local network or on the Internet as long as the other users on the network have *Cool Chat* installed on their desktop.

---

[64] Speech Synthesis: Technology for Disabled People, Edwards, A.D.N, page 6.

It can be substituted for the standard WinPopUP program (included with Windows 95/98) and runs on a TCP/IP network.

Lan Talk Pro version 2.6.1.0 is a commercial application, intended for office use to enable colleagues to communicate with each other regardless of location as long as they are on the same computer network and have the application on their desktop as well. Similar to *Cool Chat*, it enables real time conversation through a dual window showing both the history of the conversation and new messages. Its technical specifications are the same as those of Cool Chat; namely, it is able to function on Windows 98 and via a TCP/IP network.

Lan Talk Pro was created by Paul Falcon and is part of a suite of products developed by exxZero Research Lab to serve as WinPopUp replacements.

Operating System

The operating system used was Windows 98 for reasons of convenience and ease of set-up, since this operating system neither requires administration access nor special configuration.

Personal computer

800 MHz processor

256 MRAM

Sound Blaster PCI 128 soundcard

Headset microphone

An adjustable headset microphone from Emkay Innovative Products with right or left side preference was attached to the computer via the microphone and speaker plugs. Its optimal position was 10mm from the side of the subject's mouth to the tip of the microphone boom.

Figure 1 is a diagram of the network and system set-up as per the specifications.

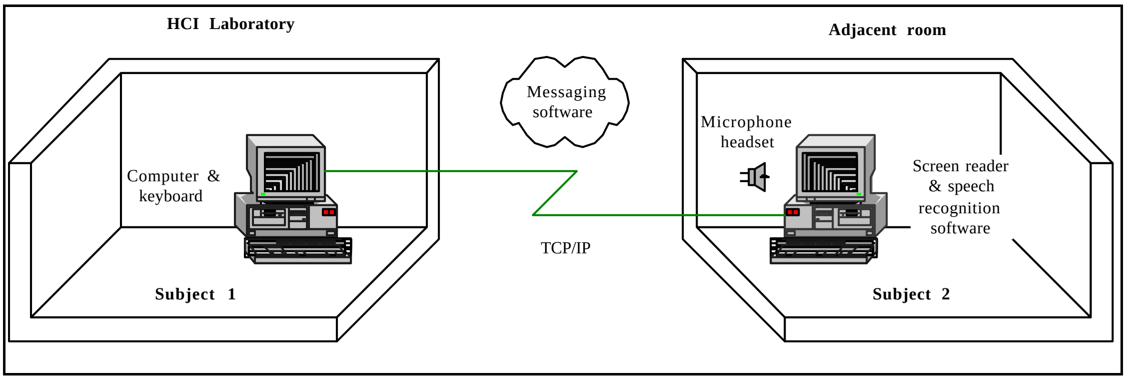

Fig. 1. An overview of the system's set-up showing the network, hardware and software

Subjects

Session one

The two subjects were first year undergraduates from the University of York's Computer Science department. Subject 1 was profoundly deaf while Subject 2 was hearing.

Both subjects were issued with individually tailored briefings (to be found in Appendix B), stating the purpose of this study, the technical details of the simulation, material for the role-plays and background information on the scenario. In order that proceedings were as natural and spontaneous as possible, the briefings were adapted to each subjects' roles, i.e. certain information was withheld from one subject, but released to the other. In this way, the subjects would possess neither prior knowledge of future events nor preconceptions about their behaviour, which would taint any resultant data.

Once the session was completed, the subjects had to fill out a debriefing questionnaire (see Appendix C), comprising five open-ended questions to gauge their immediate reactions to the system, its performance and their general thoughts and recommendations. The idea being that during the report write-up their opinions could be incorporated into the report as supporting material for any conclusions made.

Session two

Once again, owing to constraints on time, the deaf candidate was another profoundly deaf person. The major difference in this case was that the subject had been born profoundly deaf, but had recently received an implant and could 'hear'. It is beyond the scope of this study to detail the precise nature of the implant and its effects. Suffice to say, a deaf person with an implant does not hear and process sounds in a same manner as a fully hearing person, s/he is taught how to interpret sounds.

Subject 1 works as an interpreter for the deaf and is actually the interpreter for session one's Subject 1 at university. Through her job she has had similar extensive exposure to the hearing world and mainstream society as session one's Subject 1. It has also meant that she is in the unique position of be able to provide vital information on all aspects of the deaf community and being deaf, and frame it in a context that hearing people can understand.

Since she was born profoundly deaf, Subject 1 was brought up with BSL and taught to lip read. Despite having had an implant operation, 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 text telephone.

Subject 2 remained unchanged from session one.

RESULTS

SESSION ONE

Prior to the actual experiments a list of points was made in anticipation of the events expected to occur during the sessions. They were derived from other experiments and literature on this topic, and were borne in mind as the study progressed.

Just before the start of session one Subject 1 tested the speech recogniser by saying the word 'over'! He had been told that his pronunciation of the word was good and he wanted to see if the speech recogniser could transcribe the word correctly. It transcribed 'over' as 'Andrew'.

Dr. Edwards, in his guise as a high-level screen reader, sat before the terminal for Subject 2. While Subject 2 himself was located away from the terminal with no view of the screen to effect the conditions of a normal telephone conversation.

Session one was relatively quick to complete with both subjects improvising with and responding well to the system. There were no technical problems during the scenarios.

Subject 2's accent was neutral English with no regional influence; he spoke at a perceptibly normal speed. Subject 1 seemed able to follow the flow of conversation in each scenario. With the exception of minor misunderstandings – once in scenario 1 at the start and throughout scenario 3 – there was no breakdown of communication between the two interlocutors.

There was a sense that both subjects were totally focused on what was being said at that moment, rather than, as in normal conversation, thinking of what to say next. The lengthy delays and pauses between each question and response did not help this situation. At the end of session one Subject 2 commented on how frustrating it was to have to wait such a long time for each response and pointed out that outside of the laboratory anyone else would have hung up, believing that Subject 1 was perpetrating some kind of joke. He suggested the use of a 'canned' message, identifying Subject 1 as a deaf person, at the start of each conversation so that the hearing person would adapt his behaviour accordingly.

In general, the system coped better with transcribing longer sentences. Even with the shorter sentences, the system did not fail completely. One related observation was that the microphone had to be close to the mouth so as to catch all the interruptions, half-sentences and incomplete words spoken.

As each scenario progressed, it became obvious that the first two scenarios (where there was a limit to the vocabulary that could be used) were relatively easy for the system to cope with. Once scenario 3 opened, there seemed to be more mis-recognitions as the conversation moved beyond a specific situation and constraints to become free flowing. (Although, later calculations showed that session one's scenario 3 had, contrary to appearance and expectations, the second highest rate of correct transcriptions out of all the sessions at 73%.) Subject 2 became more animated, which was most definitely not captured by the system. In addition, the time delays had an adverse impact on the flow of the conversation; twice both subjects had to change the topic since the other could not understand what was being said.

In order that the accuracy of the Dragon Point and Speak could be measured, a tape recorder (Sanyo Microcassette Recorder TRC-515M) next to Subject 2 recorded the session. After the experiment the recordings of what Subject 2 said were compared to what the Dragon Point and Speak had transcribed for Subject 2. At the end of each scenario the chat log was saved as a text file for the later analysis.

SESSION TWO

In contrast to session one, this session had not been planned in advance and was relatively spontaneous. The idea was to confirm or disprove the outcomes of session one by using the same Subject 2, but changing Subject 1. Hence, this session took place once a rough evaluation of session one had been made to see if the same patterns could be noted during the second session. At that time, reading through the first set of transcripts seemed to show that scenario 3 was beyond the capability of the speech recognition software. It seemed unable to cope with anything beyond a limited vocabulary range and the delays and mis-transcriptions were excruciatingly difficult and prolonged to resolve. The decision was made that this session would not include a scenario 3 and the focus would be on scenarios 1 and 2 instead.

All specifications remained the same as per session one with two notable exceptions. Firstly, Dr. Edwards was no longer available to resurrect his role as a high-level screen reader as in the earlier session. The author assumed this function in his place. It was comparable to selecting a different voice to read out the text as is commonly done with other screen reader programs. Secondly, the setting for Dragon Point & Speak was reset to 'most accurate', during session one the 'fastest' speech recognition option had been selected.

The proceedings were no different from what occurred in session one with the exception that the delays were longer, as Subject 1 was unfamiliar with using the keyboard and was unsure of where to type in the messaging dialog box. Unfortunately, owing to a technical problem with the tape recorder, the recordings of scenarios 2a and 2b were lost and no comparison could be made between the transcripts and them. The transcript for scenarios 1a and 1b can be found in Appendix A.

As an exercise to find more evidence to support the evaluation of session one, session two was not a success. However, it did yield a wealth of information on understanding the deaf mentality and the problems confronting deaf users, which have been incorporated into this document (especially in the **Constraints of the study**).

SUMMARY

Telephony is person-to-person communication through a device. Nevertheless, at both ends of the connection, it is still a human-to-machine interaction, giving rise to issues associated with human-to-machine interface coupled with those from using natural language with a system.

Using the list prepared prior to the experiments, the transcripts from the sessions were compared and noticeable patterns and problems summarised. Based on these observations, a further list of problems was produced and segregated into two categories: those arising from poor speech recognition software or microphone quality; and inherent problems of speech recognition. Each point has supporting examples, taken from transcripts of both sessions. Note that the lines in parentheses are what were said and those in regular are their transcriptions:

Inaccuracies owing to poor speech recognition software

1. Compound and simple nouns and whole sequences of verbs and nouns were changed to names – either of people or locations.

Session	Scenario	Line	
1; 2	1a;1b	6; 3	'Rachel CV' (*read your CV*)
1	1b	2	'Mundy Tuesday' (*Monday or Tuesday*)
2	1b	8	'Lockerbie' (*one o'clock*)
2	1b	5	'the Patsy cheered' (*that would suit you*)
2	1b	20	'when Ossie' (*I'll see you*)
1	2a	11	'America Carpenter' (*I have your card number*)
1	2a	25	'at AV….archers Booker' (*if you….I'll just book*)
1	2b	37	'before Vale' (*before the film*)
1	2b	4	'in Amman' (*cinema*)
1	2b	26	'Khaled' (*card*)
1	3	33	'Brigadier' (*are you going to go*)

There is scope for misunderstanding as demonstrated during session two when Subject 1 asks (line 19, scenario 1b): "great i look forward to meeting you on Thursday 4pm, does Rachel work with you?" She is referring back to line 3 of the same scenario when the 'read your CV' is mis-transcribed as 'Rachel CV'! She read the name and assumed that it was the name of Subject 2's colleague.

2. There was replacement of certain words with similar sounding words.

Session	Scenario	Line	
1	1a	16	'wear' (*where*)
1	1b	10	'then' (*Ben*)
1	2a	31	'by' (*bye*)
1	2b	2	'below' (*hello*)
1; 2	2b; 1b	8, 35; 20	'know' (*no*)

3. The system did not note Subject 2's errors during session one. These errors ranged from repetition of words to stuttering and interruptions.

Session	Scenario	Line	
1	1a	20	'goodbye' (*goodbye said twice*)

Session	Scenario	Line	
1	3	12	‘though’ (*ttthey*)
1	3	17	‘but Bucks’ (*bbut*)

Conversely, session two saw the speech recogniser pick up Subject 2’s repetition in scenario 1b, line 10: ‘yes that's that's fine is four o'clock OK’ (*Yes, that’s that’s fine, is four o’clock okay?).* A possible explanation for this difference in performance is the altered setting used in session two, which was ‘most accurate’ (while in session one the speech recogniser was calibrated to ‘fastest’).

4. The speed of utterances sometimes resulted in gibberish being transcribed. As with point 1 later under ‘**Inherent problems of speech recognition**’, this could cause confusion and misunderstanding.
5.

Session	Scenario	Line	
1	1a	3, 4	‘yesterday and then to higher eyes ledger CV hands I mentioned interviewing a further drop’ (*Hi, I’ve read your CV and I’m interested in interviewing you*)
2	1b	12	‘our offices are in the year in assigns parking you all clear the university to think you can find that’ (*The offices are in York Science Park, which is near the university. Can you find that?)*

6. Words were compressed and substituted with single verbs/nouns.

Session	Scenario	Line	
1	1a	16	‘Askew’ (*ask at the)*
2	1a	3	‘isolated’ (*I’ve read your*)
2	1a	9	‘sorry’ (*are in*)
1	1b	2	‘future’(*for you to*)
1	2a	23	‘determine’ (*can you tell me*)
1	2a	25	‘archers’ (*I’ll just*)
1	2b	2	‘city screens in Amman’ (*City Screen Cinema*)
1	2b	2	‘propaganda’ (*I’ll put that*)

Session	Scenario	Line	
1	3	13	‘and gnome’ (*a known*)
1	3	31	‘IU’ (*are you*)

7. There was a certain pattern to the word substitution/mis-transcription, in that the system repeatedly mis-transcribed the same words.

Session	Scenario	Line	
1; 2	1a, 1b, 2a; 1b	9, 6, 5; 8	‘with’ (*would*)
1	2b	15	‘which’ (*would*)

7. Both subjects had to rely on key words in each sentence to understand the gist of the sentence,

Session	Scenario	Line	
1	2a	13	‘do you tell me a credit card number plays’ *(Can you tell me a credit card number please?*)

8. There was a lack of consistency in transcribing numbers and times. Line 5 of scenario 2a demonstrated the potential for misunderstanding when the speech recogniser mis-transcribed the time Subject 2 says: ‘with 650 be reasonable time’ (*Would 6.30 be a reasonable time?*)

   If Subject 1 had not repeated this mis-transcription and alerted Subject 2 to the mistake made, a possible consequence would have been Subject 1 purchasing the cinema tickets for a showing at a non-existent time.

9. The speech recogniser was so sensitive that it picks up the sound of Subject 2’s breathing and transcribed it as ‘and’:

Session	Scenario	Line	
2	1a	9	‘science park and to know’ (*Science Park. Do you know*)
1	2a	9	‘yes and student’ (*yes, student*)
1	2b	13	‘that’s fine and how many’ (*That’s fine. How many*)

Within certain contexts this does not represent much of a problem, but it is still a misrepresentation of information and hence a possible source of misunderstanding.

10. Rate of success:

To determine how successful the speech recognition software was, all words correctly transcribed in each scenario of sessions one and two were counted and given as a percentage of the total number of words that Subject 2 had said for that particular scenario.

Session	Scenario	How many words were correctly transcribed?	%
1 1	1a 1b	75 of a total 113 29 of a total 48	66 60
1	2a	64 of a total 102	63
1	2b	171 of a total 230	74
1	3	157 of a total 215	73
2	1a	57 of a total84	68
2	1b	77 of a total 130	59

Inherent problems of speech recognition

1. Many sentences ran into one another due to lack of punctuation, which was confusing and posed a potential source of misunderstanding. Also, other markers that qualify text when written, and are expressed by intonation when spoken were missing, such as possessives and question marks.

Session	Scenario	Line	
1	2a	20	'what is the cardholders name' (*What is the cardholder's name?*)

Subject 2 had the option of dictating punctuation, but in selecting this option, it would have contravened the principles of telephony.

2. Intonation is important in furthering understanding. In writing intonation is expressed through punctuation, however, when punctuation is absent, then the significance of a phrase can be

subverted. Line 17 of scenario 2a is a good example of such subversion of meaning: 'that's fine' (*That's fine?*)

From the recordings it is clear that that Subject 2 raised his voice to make the sentence a question. Despite the system capturing the lexicon and the syntax of the sentence correctly, the semantic was lost. This point is further underlined by the use of the same phrase 'that's fine' in a different context; in this case the phrase had the same meaning as okay, indicating agreement (line 29 of scenario 2a): 'yes that's fine' (*yes, that's fine*)

Lines 4 and 19 of scenario 3 provide another pertinent example where 'sorry?' lost its questioning aspect and became an apology, which was not the case in this context: 'sorry' (*sorry?*)

The lack of intonation recognition also resulted in loss of emphasis. Line 29 of scenario 3 shows Subject 2 being emphatic by drawing out the 'yes' at the start of his response. However, the system did not capture this and instead transcribed it as 'yet this way'.

CONCLUSION

Both Subject 1s had used Typetalk before and automatically set Typetalk as the benchmark against which to measure the performance of the system. Subject 2 had never conversed with a deaf person before, nor had he ever used Typetalk. All subjects were requested to complete a questionnaire, outlining their impressions and any suggestions for improvements. There was also an informal debriefing chat at the end of each session to gauge general reaction.

All agreed that the exchanges were time consuming as a result of the frequent misintepretations. However, according to the deaf subjects, the delays were insignificant in comparison to Typetalk. For Subject 2 they were "*disconcerting*" and "*frustrating*", and had the scenarios been for real, 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. So, the hearing person expects delays and is patient. The delays were the only major criticism from Subject 2. He also mentioned a sense of loss of control, partly caused by the delays, but mainly owing to an uncertainty as to whether Subject 1 understood what he was saying.

Overall, Subject 2 was surprised by how well the system worked with regards to speech recognition. He commented on how when it worked – that is, transcribed his words correctly – it worked, but when it went wrong, it went "*very wrong*". He believed that typing supported by instant messaging software would be as fast, and more importantly, would involve less ambiguity. Although he realised at the same time that this kind of interaction would not be strictly telephonic.

For Subject 1 from session one the overriding consideration was when to respond to Subject 2. As has been discussed earlier, prosodic features are tricky for speech recognisers to convey. Punctuation does encapsalute prosody to a certain extent, but during the sessions there was no punctuation (for Subject 2), which prompted Subject 1 to suggest using some sort of code or indicator for sentence endings. This would enable a deaf person to know precisely when her/his response was required. Something similar is already in operation with Typetalk. When a hearing speaker completes his/her sentence, s/he says, "Go ahead" or "Over to you", to indicate to the operator s/he is finished and is waiting for a response.

Apart from the concern about the timing of responses, Subject 1's only other dislike was the inaccuracy of the speech recogniser. He did however recognise that this was dependent on the software used and could be improved by installing another program. In the greater context of comprehensible and direct communication achieved through speech technology, he felt his concerns were minor and easily assuaged.

Subject 1 was overwhelming positive about the system and appreciated how communication was just between him and Subject 2. Unlike Typetalk, 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, Subject 1 said, "*speech recognition is like instant response. With Typetalk you wait for quite a while when the operator types the response*".

Another general observation made by all the subjects related to the rhythm of the thoughts: normal telephony causes speakers think of what they are about to say. In this study the subjects were focused on what was being said.

All subjects agreed that they found the system to be a "*great 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 highly unfeasible for conducting conversations and anything unstructured. Session one's Subject 1 was even more precise and stated he would use a similar system in preference to Typetalk if it were two or three times better than the study's version. His suggestion for making the system better were improved speech recognition software and for the speech recogniser to be speaker independent. Similarly, Subject 2 pinpointed the need for more efficient and faster speech recognition as well as fixed vocabulary to improve service. Notwithstanding the problems, Subject 2 conceded the system consistently met its primary objective of enabling direct communication between him and Subject 2. In light of this, he would still be happy to use the current system provided that it was in a structured situation and the vocabulary range was limited.

As for session two's Subject 1 she echoed the other Subject 1's negative as well as positive sentiments. She could offer no technical improvements, but mentioned how she appreciated the fact that she did not have to possess any particular technical knowledge or require any training to use the system. Her overall comments were more concerned with the constraints of the study and recommendations for recruiting the right sample for any further studies.

As a final note to this conclusion the most pertinent points have been summarised and, as per **Summary**, separated into those related to the speech recognition software and those referring to speech recognition in general. It is important to keep in mind that they have been derived from comments and observations made by the subjects as well as the analysis of the scenario transcripts. They should therefore not be taken out of context and do not in any way represent generalisations applying to all deaf and hearing populations.

Speech recognition software

?? A means of detecting and correcting errors, such as stuttering, incomplete sentences and interruptions is needed.

?? The system has to retain a short-term record of the most recent words. This approach could resolve the problem of scenario 2b (session one) where 'card' was mistranscribed 'cartel', 'content', and 'Khaled' within a few sentences of each other.

?? There should be a priority list of words of the same sounds, based on their frequency of occurrence in spoken language. If this had been available, it is highly doubtful that the system would have transcribed 'know' in place of 'no'. If the principle of frequent occurrence had been applied, 'no' would have been selected without a doubt.

?? The microphone quality should be improved as to ensure that a speaker could prevent his/her breathing being transcribed by the system. A control button to switch the microphone on/off could also be used to provide greater control over the microphone.

Generic speech recognition

?? Any system should not assume that users possess knowledge of speech technology or familiarity with hardware to use that system – essentially, the system should be as open-access as possible, in much the same way that text terminals are.

?? Lack of punctuation can skew a sentence's meaning and result in strings of text being compressed into a single word. In the absence of punctuation a suggestion is to have the hearing speaker pause at intervals, such as at the end of each sentence. Perhaps the pauses should be of set duration; that is, a pause to indicate the end of a sentence should last a few seconds. The duration of the pause after a word should be relative to the end-of-sentence pause. The speech recogniser would detect the pause and insert a full stop or a comma depending on the length of the pause. However, this approach is only deals with full stops and commas and does not include question marks or exclamation marks. Another solution would be to borrow from dictation programs, which allow the user to insert punctuation using abbreviations or other short cuts. However, within the specific context of telephony between a deaf and a hearing person, this approach is time consuming and relies on the hearing person knowing the abbreviations (hence, specialist training or at least some familiarity with the software). In other words, speech technology no longer enables spontaneous, direct communication with a random hearing person who may never have spoken to a deaf person before. The hearing person would need to have had some sort of prior experience.

?? The hearing speaker must not use contractions, such as 'I'll', and should use the full 'I will'. Such contractions leave the way open for a system to mis-transcribe them leading to misunderstanding. Line 24 of scenario 2a from session one is a good example of a mis-transcription arising from a contraction of the pronoun and verb: ‘archers’ (*I’ll*).

Using pronoun-verb/noun-verb contractions could also make it difficult for the speech recogniser to distinguish between a contraction and a possessive, for example, in session two, scenario 1a, line 9 and scenario 1b, line 14 ‘it is’ was mis-transcribed as ‘its’. Or between a contraction and a plural form of the noun, for example, in session 2, scenario 1b, line 14 when ‘office’s’ was transcribed as ‘offices’.

?? The hearing speaker must verify occasionally that the other party understands the flow of the conversation and has followed it correctly. In session one Subject 1 aided the flow of scenario 3 by regularly confirming information given by Subject 2. If that were not true, scenario 3 would have been resulted in even more misunderstandings than actually occurred.

?? Any hearing speaker receiving a call from a deaf party needs notification or clarification that s/he is conversing with a deaf counterparty in order that s/he measures her/her speech – either slowing it down or breaking up his/her speech into ‘digestible’ sections. Another good reason for this course of action is that the speaker realises s/he has to be patient and will be facing invariable delays during the conversation.

APPENDIX A

Transcripts of session one

All the transcripts have been coded; that is, all sentences have been numbered for easier reference. In addition, Subject 1's lines are in regular print while Subject 2's sentences have been put in bold and the correct versions (directly below the speech recogniser's output) have been given in italics.

Scenario 1a – Arranging a meeting

(1) Subject 1: good afternoon, richard

(2) Subject 1: it is ben fletcher here

(3) Subject 2: yesterday and then

(4) Subject 2: to higher eyes ledger CV hands I mentioned interviewing a further drop

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

(5) Subject 1: i'm not sure if i understand you

(6) Subject 2: woman's I have Rachel CV and I would like to interview you

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

(7) Subject 2: KD here in

Can you hear me?

(8) Subject 1: right, let me have a look at my diary

(9) Subject 2: I'm free on Wednesday the 29th August for the entire day without seeking

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

(10) Subject 1: i'm also free on that day

(11) Subject 2: okay does ten o'clock seem reasonable time

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

(12) Subject 1: what time would be appropriate for you?

(13) Subject 2: of 10 in the morning with a good for main

I think 10 in the morning would be good for me

(14) Subject 1: that's fine with me

(15) Subject 2: okay how offices are in the York assigned sparkling.

Our offices are in the York Science Park

Scenario 1a – Arranging a meeting continued

(16) Subject 2: if you Askew reception they should be able to direct human to wear an interviewer be taking place

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

(17) Subject 1: right, thank you

(18) Subject 2: okay I shall see you on Wednesday 29th August 10 o'clock in the science park

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

(19) Subject 1: ok then, thanks. bye bye

(20) Subject 2: goodbye

Goodbye. (Repeats twice)

Scenario 1b – Arranging a meeting

(1) Subject 1: good afternoon, richard. i'm afraid that i can't make the interview at the date we agreed earlier today.

(2) Subject 2: okay would be better future, on the Mundy Tuesday beforehand all later awake

Okay, would it be better for you to come on the Monday or Tuesday beforehand or later in the week?

(3) Subject 1: monday and tuesday are fully booked.

(4) Subject 2: okay how it is Friday late afternoon

How is Friday late afternoon?

(5) Subject 1: yes, i will be free on that day. same time? ten o'clock?

(6) Subject 2: actually either meeting its ten o'clock with three in the afternoon the okay

Actually I have a meeting at 10 o'clock. Would 3 in the afternoon be okay?

(7) Subject 1: sorry, i've just realised that you said late afternoon. how about 3 o'clock then?

(8) Subject 2: three o'clock is fine

3 o'clock is fine

(9) Subject 1: excellent. see ya then, Subject 2

(10) Subject 2: that by then

Goodbye, Ben

Scenario 2a – Buying cinema tickets

(1) Subject 1: good afternoon. is this the city screen cinema?

(2) Subject 2: yes how can helping

Yes, how can I help?

(3) Subject 2: ways still a

Are you still there?

(4) Subject 1: i would like to book for two student discount tickets to watch planet of the apes on thursday, preferably after six in the evening

(5) Subject 2: with 650 be reasonable time

Would 6.30 be a reasonable time?

(6) Subject 1: ten to seven, you mean?

(7) Subject 2: 630

6.30

(8) Subject 1: half past six?

(9) Subject 2: yes and student discount tickets are three pounds 50 each

Yes, student discount tickets are £3.50 each

(10) Subject 1: so, seven pounds in total?

(11) Subject 2: yes and America Carpenter plays

Yes, can I have your card number please?

(12) Subject 1: sorry?

(13) Subject 2: do you tell me a credit card number plays

Can you tell me your credit card number please?

(14) Subject 1: yes, hang on for a minute

(15) Subject 1: will a mastercard do?

(16) Subject 1: sorry?

(17) Subject 2: that's fine

That's fine?

(18) Subject 1: ok, the number is ...

(19) Subject 1: 2131-3134-4242-2586

(20) Subject 2: what is the cardholders name

What is the cardholder's name?

(21) Subject 1: it is MR F.J GILCHRIST

Scenario 2a – Buying cinema tickets continued

(22) Subject 2: okay and determine the expiry date

Okay, and can you tell me the expiry date?

(23) Subject 1: 11/03

(24) Subject 2: at AV hang on a minute archers Booker ticket

Okay, if you hang on a minute, I'll just book the ticket.

(25) Subject 1: ok

(26) Subject 2: locate two tickets are confirmed metering students ID if you want to get him with those tickets

Okay, two tickets are confirmed. You need to bring your student ID if you want to get in with those tickets.

(27) Subject 1: will NUS cards do?

(28) Subject 2: yes that's fine

Yes, that's fine.

(29) Subject 1: thanks, ok then. bye bye

(30) Subject 2: thank you by

Thank you. Goodbye.

Scenario 2b – Buying 2 cinema tickets

(1) Subject 1: hello?

(2) Subject 2: below the city screens in Amman

Hello, City Screen Cinema

(3) Subject 1: in what?

(4) Subject 2: city screens in Amman how can help you

City Screen Cinema! How can I help you?

(5) Subject 1: i've heard that there's a problem with issuing me the tickets for a film we want to watch?

(6) Subject 2: which film is secured like to watch

Which film is it that you would like to watch?

(7) Subject 1: we would like to watch the @planet of the apes@ film

(8) Subject 2: know it is quite easy to book about film and when the like to say it in

No, it is quite easy to book for that film. When would you like to see it?

(9) Subject 1: is on thursday possible? preferably after 7

(10) Subject 1: six i mean

(11) Subject 2: there's one showing at 630 and another nine o'clock

There is one showing at 6.30 and another at 9 o' clock.

(12) Subject 1: i'd like the 630 slot

(13) Subject 2: that's fine and how many people are many tickets which you like

That's fine. How many people…how many tickets would you like?

(14) Subject 1: two, student discount tickets, please

(15) Subject 2: that'll be three pounds 50 each to tell me the credit card number

That'll be £3.50 each. Could you tell me the credit card number?

(16) Subject 1: 1212-3232-4344-5666

(17) Subject 2: thank you and the name of the cartel that

Thank you and the name of the cardholder?

(18) Subject 1: MR E. CANTONA

(19) Subje ct 2: and the expiry date of the content

And the expiry date of the card?

(20) Subject 1: 13/03

(21) Subject 1: 12/03 i mean

Scenario 2b – Buying 2 cinema tickets continued

(22) **Subject 2: okay if you 81 minutes at us put down through the system that's OK propaganda the system: one minutes**

Okay, if you wait one minute I'll just put that through the system. That's okay, I'll put that through the system. Hold on one minute.

(23) Subject 1: sorry?

(24) **Subject 2: I'm sorry but the system is rejected your Khaled**

I'm sorry but the system has rejected your card.

(25) Subject 1: Khaled what?

(26) **Subject 2: the system is rejected you can't you tell me the car number and name and expiry date for more time please**

The system has rejected your card. Can you tell me the card number and name and expiry date one more time please?

(27) **Subject 2: the system has rejected your card can you give me the details one more time**

The system has rejected your card. Can you give me the details one more time?

(28) Subject 1: ok

(29) Subject 1: 1212-3232-4344-5666

(30) **Subject 2: okay**

Okay

(31) Subject 1: MR. E. CANTONA

(32) **Subject 2: and the expiry date**

And the expiry date?

(33) Subject 1: 12/03

(34) **Subject 2: one minute outright card again**

One minute. I'll try the card again.

(35) **Subject 2: know I'm sorry your card is still be rejected and you got another can't you can choose**

No, I'm sorry. Your card is still being rejected. Have you got another card you can use?

(36) Subject 1: um, i'm afraid none

(37) **Subject 2: okay that the best thing I can do is to reserve the tickets under your name and then you can collect them just before Vale**

Scenario 2b – Buying 2 cinema tickets continued

Okay, the best thing I can do is reserve the tickets under your name and then you can collect them just before the film.

(38) Subject 1: just before what?

(39) Subject 2: just before the film starts

Just before the film starts.

(40) Subject 1: ok, cheers. bye

(41) Subject 2: by an

Goodbye.

Scenario 3 – Personal chat

(1) **Subject 2: below is bent that**

Hello, is Ben there?

(2) Subject 1: Subject 2, what are you doing here? in york?

(3) **Subject 2: a than is not allotted by pretending to do resurgent**

Not a lot! I'm pretending to do research.

(4) Subject 1: sorry?

(5) **Subject 2: I'm pretending to do resurgent**

I'm pretending to do research.

(6) Subject 1: right, research you mean?

(7) **Subject 2: yes I'm being sponsored by Microsoft just the summer**

Yes, I'm being sponsored by Microsoft just for the summer.

(8) Subject 1: have they said anything about linux?

(9) **Subject 2: and no not at all than we put it on the laptop quietly but are not going to tell one**

No, not at all! We put it on the laptop quietly but we're not going to tell them.

(10) **Subject 2: everything is Windows 2000**

Everything is Windows 2000.

(11) Subject 1: is installing linux into your ipaq possible?

(12) **Subject 2: I think it is how things you can get off the Internet but though they don't you know don't seem to do any more than what Windows dozen so I haven't tried to install 1**

I think it is. There are things you can get off the Internet but tttthey don't seem to do anymore than what Windows does so I haven't tried to install one.

(13) **Subject 2: I don't think they have a console I think it in it looks like and gnome desktop**

I don't think they have a console. I think it looks like a known desktop.

(14) Subject 1: yes, that's right. i've seen some screenshots.

(15) **Subject 2: the caring article 1 furore uses a palm you've got**

Okay. Have you got one for – is it a palm you've got?

(16) Subject 1: so, what have the university built in the campus so far?

(17) **Subject 2: it doesn't look like they've got any malt may have a few months ago but Bucks they are all and working so I suppose that by the start of a new academic year they hope the shop the open**

Scenario 3 – Personal chat continued

It doesn't look like they've built any more than a few months ago bbbut they are working so I suppose that by the start of the new academic year they hope the shop will be open.

(18) Subject 1: any scotch malt?

(19) Subject 2: sorry

Sorry?

(20) Subject 1: sorry, could you please say that again

(21) Subject 2: never mind 20 data in dear

Never mind. When are you going to India?

(22) Subject 1: again, i'm not clear

(23) Subject 2: when they you going to India

When are you going to India?

(24) Subject 1: probably soon, have to be back by november for my 21st party!

(25) Subject 2: okay I did he starts

Okay, when does uni start?

(26) Subject 1: sorry?

(27) Subject 2: British University start

When does university start?

(28) Subject 1: do you mean, when british universities start their first terms?

(29) Subject 2: yet this way you coming back for the second year

Yes (drawn out yes), are you coming back for the second year?

(30) Subject 1: ah, i've got something to tell you. i'm taking a year of absence ...

(31) Subject 2: okay but IU continuing this course when you get back

Oh, okay, but are you continuing this course when you get back?

(32) Subject 1: to heal my RSI, which has been healing nicely over the summer

(33) Subject 2: to okay Brigadier travelling more

Oh, okay, are you going to go travelling more?

(34) Subject 1: that's right, next easter

(35) Subject 1: that's correct

(36) Subject 2: where are you doing together in

Where are you going to go?

(37) Subject 1: sorry, rich. must run!

(38) Subject 2: barricaded by then

Scenario 3 – Personal chat continued

Okay, bye then?

(39) Subject 1: bye

APPENDIX A

Transcripts of session two

Scenario 1a – Arranging a meeting

(1) Subject 2: other Mackie each richard manly here

Hello Maggie, this is Richard Manley here.

(2) Subject 1: hi maggie here how are you

(3) Subject 2: hi-fi thank you isolated CT and I would like to interview you

I'm fine, thank you. I've read your CV and I would like to interview you.

(4) Subject 1: when would be a suitable time for you

(5) Subject 2: use next Wednesday reasonable I've got nothing all-day

Is next Wednesday reasonable? I've got nothing on all day.

(6) Subject 1: next wednesday fine, better in afternoon cos i hate mornings!

(7) Subject 2: okay how about three o'clock on Wednesday afternoon

Okay, how about three o'clock on Wednesday afternoon?

(8) Subject 1: ok where would you like to meet, remember I am not familiar with York

(9) Subject 2: our offices sorry York science park and to know that since its near the University

Our offices are in York Science Park. Do you know where it is? It's near the university.

(10) Subject 1: ok ive been to university a few times so will find im sure

(11) Subject 2: if you go to the reception in the science park there'll be able to direct key to the meeting

If you go to the reception in the Science Park, they'll be able to direct you to the meeting.

(12) Subject 1: that sounds fine.see you wednesday at 3pm

(13) Subject 2: okay good-bye

Okay, goodbye

(14) Subject 1: thanks for calling

Scenario 1b – Arranging a meeting

(1) Subject 2: hello Maggie this is Richard Manley

Hello Maggie, this is Richard Manley.

(2) Subject 1: hello Richard, how help you

(3) Subject 2: -Rachel CT and I would like to interview you

I've read your CV and I would like to interview you.

(4) Subject 1: My CV must have impressed you,do you have a time and date in mind?

(5) Subject 2: I'm free all the Wednesdays for any time Wednesday the Patsy cheered

I'm free all of Wednesday. Is there any time Wednesday that would suit you?

(6) Subject 2: we still MIT

Are you still there Maggie?

(7) Subject 1: sorry just making some notes...sorrry Im not free on Wednesday could we make it Friday?

(8) Subject 2: Fridays flying in the afternoon with chronic Lockerbie or lights

Friday's fine in the afternoon. Would one o'clock be all right?

(9) Subject 1: Could we meet after 3.15 on Thursday afternoon?

(10) Subject 2: yes that's that's fine is four o'clock OK

Yes, that's (said twice) fine, is four o'clock okay?

(11) Subject 1: Great, where do i need to meet you?

(12) Subject 2: our offices are in the year in assigns parking you all clear the university to think you can find that

Our offices are in, near by the Science Park in York, near the university. Do you think you can find that?

(13) Subject 1: Your last message does not make sense,please retry

(14) Subject 2: the offices in the York assigns pocketed you find that its near the university

The office's in York Science Park. Can you find that? It's near the university.

(15) Subject 1: What is assigns?

(16) Subject 2: the offices starlings York assigns spark which is near the University can you find that

The offices are in York Science Park, which is near the university. Can you find that?

(17) Subject 1: I think I understand!! Are there plenty of car parkimg space

(18) Subject 2: yes

Yes.

Scenario 1b – Arranging a meeting continued

(19) Subject 1: great i look forward to meeting you on Thursday 4pm, does Rachel work with you?

(20) Subject 2: know when Ossie on Thursday

No, I'll see you on Thursday.

(21) Subject 1: again your last sentence is confused

(22) Subject 2: I shall see you on Thursday

I shall see you on Thursday.

(23) Subject 1: thats better.see you then bye

(24) Subject 2: goodbye

Goodbye.

APPENDIX B

BRIEFING FOR SUBJECT 1

Purpose of this study:

To observe the human interaction factors in using a specific telephony system. The telephony system facilitates direct communication between a deaf person and a hearing person by using speech recognition and synthesis technology.

Current telephony systems that enable a deaf person to 'speak' with a hearing person are inconvenient and non-confidential since they involve an operator/intermediary. The solution would be as stated above – a speaker independent recognition system. However, such a system is not available. In the meanwhile, the question is how to make this system reliable when the technology is fully available.

The study aims to provide some clues by simulating such a system using a speaker dependent speech recogniser, chat software and speech synthesiser with the necessary hardware, and then uncovering the human interaction factors involved in its usage.

Details of the study:

A network of two computers has been set up in the HCI laboratory and the adjacent room. Computer two has speech recognition software – Dragon Point & Speak-, a screen reader – Dr. Edwards – and a headset microphone. Both computers are linked by chat software – Cool Chat/Lan Talk Pro.

The deaf subject – Subject 1 - will use computer one and Subject 2 – the hearing subject – will use computer two.

Subjects 1 and 2 will role-play and act out two scenarios and a further two, which are variations on the original two. Finally, there is a chance for the subjects to have a personal conversation. The scenarios will last a few minutes and should follow the flow of a typical telephone conversation.

Plan of action:

- ?? Introduction – any questions or explanations of the system usage will be dealt with during this time.
- ?? Role-play – scenarios; a recording will be made of the session.
- ?? Debriefing – a questionnaire will be completed; general comments will be gathered.

Role-play:

Scenario 1 a – Arranging a meeting

You are looking for a job and have submitted your CV to Subject 2's company. He likes it so much that he wishes to meet you face-to-face so he contacts you to arrange a lunchtime interview.

Monday 27th August 2001	**Thursday 30th August 2001**
9.15am-5.15pm: lectures	9.15am-11.15am: C programming lecture 12.15pm-3.15pm: Football training
Tuesday 28th August 2001	**Friday 31st August 2001**
10.15am-4.15pm: Electronics workshop	Travelling to London to see some friends
Wednesday 29th August 2001	**Saturday 1st September 2001**

Using the diary above, try to arrange a date and a time with him that are suitable for both you and him. He will suggest the location. Confirm with him your availability subject to your diary.

The emphasis is on being natural, as if you were talking on a real telephone. Improvise if you wish to since there are no rules regarding length or exact content of the call.

Scenario 1b – Arranging a meeting

Monday 27th August 2001	**Thursday 30th August 2001**
9.15am-5.15pm: lectures	9.15am-11.15am: C programming lecture 12.15pm-3.15pm: Football training
Tuesday 28th August 2001	**Friday 31st August 2001**
10.15am-4.15pm: Electronics workshop	No trip to London – cancelled!
Wednesday 29th August 2001	**Saturday 1st September 2001**
11.15am-2.15pm: lunch with visiting parents	

As with scenario 1a, the parameters are the same although in this scenario there will be a difficulty, which will involve you having to talk with Subject 2 for a longer time period than in scenario 1a.

You cannot make the time that Subject 2 arranges for the lunchtime interview. It is up to you to determine which date and time would be more appropriate based on your diary.

Scenario 2a – Purchasing two cinema tickets

You wish to go to City Screen cinema in York to see Planet of the Apes with a friend. You need to buy two student discount tickets and you wish to go on Thursday, preferably after six in the evening.

You have to find out the price, the time of the show and whether there are any student discounts. You also pay for the tickets by giving your:

?? Card number
?? Name of card holder
?? Expiry date

Scenario 2b – Purchasing two cinema tickets

As with scenario 2a, the context is the same with the exception that this time round there is a problem. Subject 2 will tell you what the problem is and you have to resolve it as you see fit.

Scenario 3 – Getting to know Subject 2 better

This is not a scenario, more a chat with Subject 2 to find out some details about him that you could relay back to me – for example, what food he likes etc. About five pieces of information should be enough!

This scenario should be more free flowing and relaxed than the others, simply because you are chatting with Subject 2 as Subject 2 and not playing a specific role. You can talk about anything; any topic that catches your fancy as you would with any personal friend of yours.

BRIEFING FOR SUBJECT 2

Purpose of this study:

To observe the human interaction factors in using a specific telephony system. The telephony system facilitates direct communication between a deaf person and a hearing person by using speech recognition and synthesis technology.

Current telephony systems that enable a deaf person to 'speak' with a hearing person are inconvenient and non-confidential since they involve an operator/intermediary. The solution would be as stated above – a speaker independent recognition system. However, such a system is not available. In the meanwhile, the question is how to make this system reliable when the technology is fully available.

The study aims to provide some clues by simulating such a system using a speaker dependent speech recogniser, chat software and speech synthesiser with the necessary hardware, and then uncovering the human interaction factors involved in its usage.

Details of the study:

A network of two computers has been set up in the HCI laboratory and the adjacent room. Computer two has speech recognition software – Dragon Point & Speak-, a screen reader – Dr. Edwards – and a headset microphone. Both computers are linked by chat software – Cool Chat/Lan Talk Pro.

The deaf subject – Subject 1 - will use computer one and Subject 2 – the hearing subject – will use computer two.

Subjects 1 and 2 will role -play and act out two scenarios and a further two, which are variations on the original two. Finally, there is a chance for the subjects to have a personal conversation. The scenarios will last a few minutes and should follow the flow of a typical telephone conversation.

Plan of action:

?? Introduction – any questions or explanations of the system usage will be dealt with during this time.

?? Role-play – scenarios; a recording will be made of the session.

?? Debriefing – a questionnaire will be completed; general comments will be gathered.

Role-play:

Scenario 1 – Arranging a meeting

You are a prospective employer. Subject 2 is looking for work and has sent you his/her CV. You like it enough to want to meet him/her face-to-face so you call him/her to arrange an interview.

Monday 27th August 2001	**Thursday 30th August 2001**
12pm: lunch with client	11am-2pm: sales meeting
Tuesday 28th August 2001	**Friday 31st August 2001**
5pm: conference call with New York office	10.30am: weekly round-up
Wednesday 29th August 2001	**Saturday 1st September 2001**

Using the diary above, arrange a time with him/her that suits both you and him/her. (Hint: Wednesday 29th August is free of any engagements.)

The call should be start with a general introduction followed by the purpose of your call then the discussion regarding the date, time and location of the meeting. (Hint: the company offices are in York Science Park.)

The emphasis is on being natural, as if you were talking on a real telephone. Improvise if you wish to since there are no rules regarding length or content of the call.

Scenario 1b – Arranging a meeting

Monday 27th August 2001	**Thursday 30th August 2001**
12pm: lunch with client	11am-2pm: sales meeting
Tuesday 28th August 2001	**Friday 31st August 2001**
5pm: conference call with New York office	10.30am: weekly round-up
Wednesday 29th August 2001	**Saturday 1st September 2001**

As with scenario 1a, the parameters are the same although in this scenario there will be a difficulty, which will involve you having to talk to Subject 1 for a longer time period than in scenario 1a. No clue has been given as to the nature of this difficulty in order that your responses are as spontaneous and natural as possible!

Scenario 2a – Purchasing two cinema tickets

You are the manager of the City Screen cinema in York. Subject 1 wishes to purchase two cinema tickets. Using the timetable of films below, help him/her select the right show time.

Monday 27th August 2001	**Thursday 30th August 2001**
Planet of the Apes (12): 1.30, 4.00, 6.30, 9.00 Lucky Break: 2.30, 4.30, 6.40, 8.40 Intimacy (18): times tbc Help I'm a fish(U): 2.15, 4.15	Planet of the Apes (12): 1.30, 4.00, 6.30, 9.00 Lucky Break: 2.30, 4.30, 6.40, 8.40 Intimacy (18): times tbc Help I'm a fish(U): 2.15, 4.15
Tuesday 28th August 2001	**Friday 31st August 2001**
Planet of the Apes (12): 1.30, 4.00, 6.30, 9.00 Lucky Break: 2.30, 4.30, 6.40, 8.40 Intimacy (18): times tbc Help I'm a fish(U): 2.15, 4.15	Planet of the Apes (12): 1.30, 4.00, 6.30, 9.00 Lucky Break: 2.30, 4.30, 6.40, 8.40 Intimacy (18): times tbc Help I'm a fish(U): 2.15, 4.15 Breakfast at Tiffany's (PG): 6.15 Film TBC: times tbc Late Nights: Raiders of the Lost Ark (PG): 10.45 The Pillow Book (18): 10.40
Wednesday 29th August 2001	**Saturday 1st September 2001**
Planet of the Apes (12): 1.30, 4.00, 6.30, 9.00 Lucky Break: 2.30, 4.30, 6.40, 8.40 Intimacy (18): times tbc Help I'm a fish(U): 2.15, 4.15	Kids Club: Dunston Checks In (PG): 11.45 Planet of the Apes (12): 1.30, 4.00, 6.30, 9.00 Lucky Break: 2.30, 4.30, 6.40, 8.40 Intimacy (18): times tbc Help I'm a fish(U): 2.15, 4.15 Breakfast at Tiffany's (PG): 6.15 Film TBC: times tbc Late Nights: Raiders of the Lost Ark (PG): 10.45 The Pillow Book (18): 10.40

You give information on times and dates of shows and the price of tickets.

Price of normal ticket: £4.00 (except on the weekends when it is £5.00)

Students' discount: £3.50 with student id (except on the weekends when there are no student discounts)

You also accept payment of tickets over the telephone by asking for:

?? Card number

?? Name of card holder

?? Expiry date

Scenario 2b – Purchasing two cinema tickets

As with scenario 2a, the context is the same with the exception that this time round Subject 1's card is not accepted. You have to inform him/her that his/her card has been rejected. Resolve the situation as you see best! (Hint: he could use another card or another method of payment.)

Scenario 3 – Getting to know Subject 1 better

This is not a scenario, more a chat with Subject 1 to find out some details about him/her that you could relay back to me – for example, find out if he/she has a cat etc. About five pieces of information should be enough!

This scenario should be more free flowing and relaxed than the others, simply because you are chatting with Subject 1 as Subject 1 and not playing a specific role. You can talk about anything; any topic that catches your fancy as you would with any personal friend of yours.

APPENDIX C

DEBRIEFING QUESTIONNAIRE

What did you notice most when using this system?

What did you find difficult when using this system?

What did you like about the system?

What would you change to make usage easier/better?

Would you ever use this system?

APPENDIX D

CURRENT TELEPHONY SERVICES

Europe saw its first mobile text telephony service in 1998 when Europolitan and Telesta, Swedish telecommunication companies, launched Mobile Text Telephony as a special mobile subscription for the deaf. Mobile text telephony enables both sender and receiver to have text displayed on their screens as it is being typed, and permits communication with fixed text telephones as well[65].

Video telephony is of most use to those deaf people with sign language as their mother tongue. For other users especially those who become deaf later in life, this technology does not provide any obvious benefit other than as a nice additional function to text telephony systems. Furthermore, its limitations prevent its application to the mobile network (unlike simple text telephony). The general consensus seems to be that video telephony should be one of the modes on a text telephony system rather than a stand-alone medium of communication. This is termed Total Conversation where users have a choice and can switch from one medium to another with ease.

65 Press release from Europolitan at http://www.europolitan.se/1015.euro

REFERENCES

Ainsworth, W.A., *Speech Recognition by Machine*, 1988, Peter Peregrinus Ltd.

Blenkhorn, Paul, *Producing a Text-to-Speech Synthesizer for Use by Blind People,* from *Extra-ordinary Human Computer Interaction Interfaces for Users with Disabilities,* edited by Edwards, Alistair D. N., 1995, Cambridge University Press

Bernsen, Niels Ole, Dybkjær, Hans, Dybkjær, Laila, *Designing Interactive Speech Systems: From First Ideas to User Testing*, 1998, Spring Verlag

Cater, John P., *Electronically Speaking: Computer Speech Generation*, 1983, Howard W. Sams & Co., Inc.

Edwards, Alistair D. N., *Speech Synthesis: Technology for Disabled People*, 1991, Chapman.

Holmes, J.N., *Speech Synthesis and Recognition*, 1988, Van Nostrand Reinhold (UK) Co. Ltd.

Howey, Kate, *Technology for People with Special Needs: HCI Design Issues*, *Extra-ordinary Human Computer Interaction Interfaces for Users with Disabilities*, edited by Edwards, Alistair D. N., 1995, Cambridge University Press

Longuet-Higgins, Christopher, *Tones of Voice: The Role of Intonation in Computer Speech Understanding*, pages 293-301 of *Computer Speech Processing*, edited by Fallside, Frank, Woods, William A., 1985, Prentice-Hall International (UK) Ltd.

Mankoff, Jennifer, *Speech Recognition – where are we, and where will we go*? She was a Georgia Tech College of Computing PhD student, now working at Group for User Interface Research, University of California Berkeley. Her speech is available at http://www.cc.gatech.edu/people/home/jmankoff/speech-writeup.html

Newell, A. F., Arnott, J.L., Dye, R, *A Full Speed Speech Simulation of Speech Recognition Machine*, Proceedings of the European Conference on Speech Technology, Edinburgh, 1987, Vol. 2, pages 410-413.

Newell, A. F., Arnott, J.L., Carter, K., Cruickshank, G., *Listening typewriter simulation studies*, International Journal of Man-Machine Studies, 1990, Vol. 33, pages 1-19.

Paul, Peter V., Quigley, Stephen P., *Language and Deafness*, 1984, Singular Press Inc. Quotes were taken from *Types of Deafness* at http://home.inreach.com/torsi/typesofD.html

Pierrel, J.M., Aspects of Man-Machine Voice Dialog, pages 249-273 of *Fundamentals in Computer Understanding Speech and Vision*, edited by Haton, J.-P., 1987, Cambridge University Press.

Pinon, Carla D., Technology evaluation (speech recognition), Stanford University, available at http://www.stanford.edu/~cpinon/speech.html

Sperandio, J.C., Scapin, D.L., *Ergonomic Aspects of Man-Machine Communications*, pages 79-90 of *Fundamentals in Computer Understanding Speech and Vision*, edited by Haton, J.-P., 1987, Cambridge University Press.

Yannakoudakis, E.J., Hutton, P.J., *Speech Synthesis and Recognition Systems*, 1987, Ellis Howard Ltd.

Details (only in Swedish) on the Nordic Forum on Telecommunications and Disability can be found at the following website: http://www.nsh.se/. Another site with information in English is http://www.design-for-all.org/magazine3/main_we_have_the_key.html. Data used in this report came from http://www.stakes.fi/cost219/Texttelephony.htm

Information on the UMTIDUMPTI project can be found at:
http://www.infowin.org/ACTS/RUS/PROJECTS/ac027.htm or
http://www.een.bris.ac.uk/UMPTIDUMPTI/data/deliverables/d21/Texttel3.htm

Europolitan AB, a Swedish telecommunications company, has a website
http://www.europolitan.se/001.euro

1st Voice, a Dragon Systems partner and reseller, can be found at:
http://www.1stvoice.com/users.html

Deafness and Mental Health, Mind Publications, available at
http://www.mind.org.uk/information/factsheets/D/deafness/Deafness_and_Mental_Health.asp

The Leeds Society for Deaf and Blind People: http://www.hipleeds.org.uk/leeds.htm

Iowa Department for the Blind, *How a screen reader works*? :
http://www.blind.state.ia.us/access/how.htm