THE UNIVERSITY OF HULL

Augmentative Communication Device Design, Implementation and Evaluation

being a Thesis submitted for the Degree of

Doctor of Philosophy

in the University of Hull

by

Richard Robert Plant, B.Sc. (Hull)

April 1996

Summary of Thesis submitted for PhD degree

by Richard Robert Plant

on

Augmentative Communication Device Design, Implementation and Evaluation

The ultimate aim of this thesis was to design and implement an advanced software based Augmentative Communication Device (ACD), or Voice Output Communication Aid (VOCA), for non-vocal Learning Disabled individuals by applying current psychological models, theories, and experimental techniques. By taking account of potential user's cognitive and linguistic abilities a symbol based device (Easy Speaker) was produced which outputs naturalistic digitised human speech and sound and makes use of a photorealistic symbol set. In order to increase the size of the available symbol set a hypermedia style dynamic screen approach was employed. The relevance of the hypermedia metaphor in relation to models of knowledge representation and language processing was explored.

Laboratory based studies suggested that potential user's could learn to productively operate the software, became faster and more efficient over time when performing set conversational tasks. Studies with unimpaired individuals supported the notion that digitised speech was less cognitively demanding to decode, or listen to.

With highly portable, touch based, PC compatible systems beginning to appear it is hoped that the otherwise silent will be able to use the software as their primary means of communication with the speaking world. Extensive field trials over a six month period with a prototype device and in collaboration with user's caregivers strongly suggested this might be the case.

Off-device improvements were also noted suggesting that Easy Speaker, or similar software has the potential to be used as a communication training tool. Such training would be likely to improve overall communicative effectiveness.

To conclude, a model for successful ACD development was proposed.

ACKNOWLEDGEMENTS

Throughout the last four years my wife and daughter have supported me 'come hell or high water' and this seems scant enough tribute for all the sacrifices they have made. However, I hope this tome, which is dedicated to them will serve as some recompense.

To my supervisor, Dr Andrew Rostron, who has acted not only as academic mentor but who has become a close personal friend. During the last four years he has aided my cause both academically and personally and to him I am greatly indebted.

My close friends, who unbeknown to them have provided many welcome breathing spaces from academia and who will hopefully continue to do so long into the next millennium.

Despite lack of formal funding I have been able to persevere over the years due to the generosity of three organisations and groups of people, without contribution from which fate may have taken a different turn. Firstly I would like to thank the University of Hull for awarding me a bursary which covered my postgraduate tuition fees. Secondly, the Psychology Dept in Hull who have not only provided the resources to help me complete my work but also for the ad hoc employment that has enabled us as a family survive. Finally, my thanks go to the charitable organisation who provided the minimum funds required to keep me ticking over financially between odd jobs.

I am grateful to North Lindsey College, Scunthorpe; Rawcliffe Hall Hospital, Goole; The Clinical Psychology Dept., Scunthorpe General Hospital, Scunthorpe; and the Sheffield Consulting & Clinical Psychologists, Learning Disabilities Service, Northern General Hospital, Sheffield; all of whom provided me with support and introduced me to individuals who might benefit from me working with them. I am especially indebted to Ms. Kath Scales (North Lindsey College), Dr. Barry Hart, and Dr. Shripati Upadyaya (SCCP, Northern General Hospital, Sheffield). My work was aided immeasurably by the hospitals mentioned, all of whom allowed me to work for them on and off as an Assistant Clinical Psychologist throughout my postgraduate studies.

My thanks and thoughts must finally go out to the four Learning Disabled individuals who took part in field trials, and who helped me to realise how fortunate I am and what real adversity means.

TABLE OF CONTENTS

PREFACE

Currently the field Alternative and Augmentative Communication (AAC) is constituted from an amalgam of specialist areas which include amongst others: Psychology, Communication Science, Speech Therapy, Medical Physics, Electronic Engineering and Computer Science. This is a consequence of the nature of the problem, namely to enable the non-vocal to communicate. Non-vocal individuals typically have many other needs in addition to their need to communicate with the speaking world. This fact tends to blur the lines of traditional field boundaries. In terms of AAC in relation to electronic or Augmentative Communication Devices (ACD) no one field has exclusive reign. Unfortunately this has meant that many ACD's are utilitarian and commonly address the needs of certain groups directly with the less able being excluded. Due to the utilitarian nature of many ACD's this has meant that little objective research has been done with certain groups; with the Learning Disabled being the largest.

One of the problems for research in AAC has been which theories should be tested, what research methodologies should be used, and ultimately how findings should be applied. Psychology is perhaps the only field whose theories, research methodologies and record of practical application can be directly targeted toward the problem. Within this thesis many psychological theories have been applied in order to design, implement and monitor, or evaluate, a software based ACD for the non-vocal Learning Disabled population. Psychological research methodology has allowed objective collection of data via automated means and by application of standard clinical inventories. Culmination of the data has enabled theories to be confirmed or rejected and has suggested future avenues for investigation.

Despite Psychology providing a framework for research classical techniques cannot be directly applied due to the nature of the problem and the individuals involved. For example, one cannot ethically with such a vulnerable group manipulate variables whilst observing the result in order to confirm or refute a hypothesis. Neither can one deprive a control group of the opportunity to benefit from such an ACD; not withstanding the difficulty of matching the abilities and handicaps of individuals anyway. Nor can one withdraw equipment or support provision in order to observe the impact. Generally working with such groups has been shied away from as typical techniques such as protocol analysis can't be applied as individuals lack the insight required. Simply, they lack the ability to tell you verbosely what they are thinking at a given time.

Research however needs to be carried out, theories formulated, tested, modified and applied. Just because standard techniques cannot be used directly doesn't mean that well controlled and meaningful research cannot be conducted. The view taken by the author is that an action research paradigm is the only way forward in such difficult areas. Whereby, a need is addressed and the research follows in parallel, as opposed to the classical model where research precedes application. To gain the trust and support of other professionals and carers in the area such an approach was seen as the only way forward.

In short the ultimate aim of this thesis has been to design and implement an advanced software based Augmentative Communication Device (ACD), or Voice Output Communication Aid (VOCA), for non-vocal Learning Disabled individuals by applying current psychological models and techniques. By taking account of potential user's cognitive and linguistic abilities a symbol based device (Easy Speaker) has been produced which outputs naturalistic digitised human speech and sound. In order to increase the size of the available symbol set a hypermedia style dynamic screen approach has been employed. With highly portable, touch based, PC compatible systems beginning to appear it is hoped that the otherwise silent will be able to use the software as their primary means of communication with the speaking world. Extensive field trials over a six month period with a prototype device has strongly suggested this might be the case. Such success in the field was only accomplished after years of research and development and by application of the rationale outlined below.

With the clear goal of producing a system which is potentially usable by even the severely Learning Disabled non-vocal, a review of abilities of the population in question was first undertaken. Within this review cognitive and linguistic abilities were considered alongside levels of unaided communicative effectiveness. The implications for ACD intervention were considered.

Before attempting to design and develop a 'new' kind of system a review of commercial hardware and software based ACD's was undertaken. Common forms of speech output and choice of symbol set were considered. Advanced approaches and the need for further development were also considered.

In order to allow users access to a large vocabulary of symbols it was decided to employ a dynamic screen approach based around a hypermedia metaphor. The relevance of the hypermedia metaphor in relation to models of knowledge representation and language processing was explored. As was the implication of ACD usage based upon such an access metaphor.

The development of the actual software was considered from an HCI standpoint in relation to both caregivers and users alike. The importance of screen layout and models of conceptual understanding were explored. Central to the notion of Easy Speaker usage was tracking users actions and output whilst using the device productively. The nature of these recording techniques was explained. The evolutionary processes Easy Speaker underwent were considered as various software revisions were written.

It is essential that care givers and health care professionals can maintain and set up hierarchical symbol vocabularies for real users easily. In order to evaluate the success of the interface for those who might maintain symbol vocabularies for users a standard Questionnaire for User Interface Satisfaction was given to naive maintainers and then again when experienced. The attitudes of maintainers to the creation and maintenance of hierarchies was favourable, and above expected levels.

The quality of speech output from any communication aid is crucial and the debate continues to centre around whether to use digitised or synthesised speech. Research has suggested that synthesised speech is more cognitively demanding to decode, or listen to. With the central aim of producing a low demand ACD it was proposed from both theoretical and practical standpoints that naturalistic digitised speech should be used. This is even more relevant as it is suggested that

ACD users also listen to and reprocess their own *spoken* output and so can affect them adversely too. The level of processing overhead was confirmed experimentally by investigating the cognitive load incurred by synthesised speech over digitised, when obeying increasingly complex instructions.

In order to evaluate the success of Easy Speaker in its role as communication aid two bespoke inventories needed to be constructed. The first, the Easy Speaker Assessment Tasks (ESAT), measured users raw performance on-device whilst generating set tasks over a three week period. All task data was collected by the Automated Response Tracking System (ARTS) as users completed the tasks. Measures focused on utterance construction speed, efficiency and estimated number of words per minute output. All were recorded fully automatically and were highly accurate. The second, the Easy Speaker Performance Monitoring (ESPM) questionnaire, was concerned with off-device improvements in language and cognition over the longer term. Immediate caregivers were asked to rate improvements in behavioural measures of each users adaptive behaviour pre-intervention and after six months of use. All items were taken from the Adaptive Behaviour Scales and the Functional Performance Record 16, and might be collectively considered as a measure of each users communicative effectiveness within their everyday environment.

With the availability of objective test materials which assessed raw performance users were tested on set tasks over a three week period (ESAT). During this three week period three of the four users improved significantly on all performance measures. This supported that the notion of a low demand hypermedia style interface using photorealistic symbols might be correct for such a user group. Furthermore when ARTS data was analysed in terms of productive quality it was found that all users had generated meaningful multi-symbol utterances with both the author and their immediate caregivers. Performance measures and productive use suggest that Easy Speaker could be used as a primary communication channel by the non-vocal Learning Disabled population.

As the current study group were borderline vocal and moderately Learning Disabled off-device measures of improvements in language and cognition were taken using recognised behavioural inventories (ESPM). After six months of Easy Speaker intervention all users improved significantly on all measures of language and cognition as rated by two of each users immediate caregivers over their previous baseline ratings. Such behavioural measures can be thought of as collectively measuring communicative effectiveness.

To conclude, a model for successful ACD development was proposed which followed a circular pattern and encompassed, (1) The theoretical stage, (2) The developmental stage and (3) The evaluative stage. It was suggested that all three stages should take place within a Framework of existing, new, and cutting-edge technologies. It was this model that Easy Speaker development followed. Future trends were considered under the headings of hardware and software advances. It was suggested that hardware advances in standard PC compatible technologies could be left to evolve naturally due to office and workplace demands by unimpaired computer users. That is, shrinking size, decreasing cost, increasing computational power and multimedia features as standard. It was suggested that ACD's should be constructed wholly in software so that they might make use of such industry standard platforms. Taking account of emerging technologies a prototype of the next version of Easy Speaker was outlined. The shortcomings of the current version, as highlighted by the QUIS study, were resolved. Maintainers could now maintain user vocabularies, or hierarchies, under an object orientated interface simply by dragging and dropping symbols onto user screens. In addition, the new prototype allowed for dynamically scaleable symbols, variable symbol density, single switch and touch screen input, and the ability to use additional media such as music, animations and even video clips. With design of the latest version concentrating on both the user and caregiver it is hoped that time and resource pressured environments, such as the NHS, will consider Easy Speaker as a choice for the severely Learning Disabled non-vocal. To this end future research directions were outlined.

CHAPTER 1: Priorities in design, implementation and monitoring of symbol based Augmentative Communication Devices

Demographic research suggests that between 8 to 12 people per 1000 are in some way profoundly handicapped in their spoken language usage. Within the Learning Disabled (LD) population this figure rises to approximately 23 to 25 per 100. Such impoverishment means that many individuals may be unable to meet their daily communication needs.

In a bid to alleviate such difficulties many Augmentative and Alternative Communication (AAC) strategies have been employed, ranging from sign language to complex computer based speech prostheses. Although such AAC strategies may facilitate communication within the otherwise more able general population, potential LD users are likely to posses other handicaps which preclude such approaches. These can encompass, intellectual and cognitive impairments, physical disability, deafness and blindness amongst others.

With restricted abilities simpler AAC's have replaced generally available alternatives. Simpler signing systems such as Makaton has replaced sign language and symbol based communication systems have been introduced. Graphically based symbol systems such as, Bliss and Rebus, have been taught to many individuals as a replacement or supplant to spoken language. Such individuals can communicate with others by sharing these symbols which have a common meaning to both *speaker* and *listener*.

With the advent of modern computer systems symbol based alternatives have mutated into computer based Augmentative Communication Devices (ACD), or Voice Output Communication Aids (VOCA). Such evolution allows language impaired individuals to interact with the normal speaking population through a symbol set which can be translated into artificial speech by the device and *spoken* to the listener.

Such ACD's have generally been 'need led' rather that research based. The 'if it works, use it!' mentality has meant that many devices are overly complex, inefficient in use, and as such accessible only be the most able of the LD population. However as 'needs dictate' such devices continue to be produced and used.

The goal of the thesis and this chapter in particular is to identify from both theoretical and research based perspectives what features an 'idealised' ACD for such a group might possess and the impact it might make on their lives. The result will be an 'idealised' software based ACD, called "Easy Speaker", which has been field trialed, and of measurable benefit to users.

INTRODUCTION

With some 23% of all LD individuals experiencing quite profound difficulties with regard to spoken language (Gibbs and Cooper 1989) many AAC strategies have been implemented in order to elevate natural abilities. These have included physical sign based systems such as American and British Sign Language (ASL and BSL) amongst others. For those who don't possess the intellectual or physical abilities to use standard signing systems simpler alternatives have been taught, e.g. Makaton. However for many individuals with profound language difficulties they cannot make full use of even simple signing systems. For this group symbol based systems provide an alternative.

Symbol based systems are typically aimed at the less able and generally provide for more functional types of discourse derived from a limited symbol set. Typically these have consisted of semi-gaphical symbols such as Bliss or Rebus which are superimposed on to a physical grid or form-board. The *speaker* would communicate by pointing to a number of symbols in turn. Because such a symbol set has a shared meaning, both *listener* and *speaker* share the same understanding of the message. However, the speaker cannot communicate via this system with normal language users as they don't share a similar frame of reference. Instead they use spoken language.

In addition because only a limited set of symbols can be displayed on a finite size grid any communicative interaction is limited by the size of the vocabulary available. One can temporally circumvent this by increasing the symbol density or by making the grid larger. Both of which increase the demands placed on the *speaker*.

With the advent of low cost, high power, readily available computer systems such simple, but effective, communication tools have in the main been recreated as electronic form-boards or more recently 'virtually' on the computer screen. Both can offer the ability to turn symbol input into synthesised or digitised speech so that users can interact easily, both with other like individuals and with members of the normal speaking population. Computer based variants can offer dynamic displays that enable 'pages' of symbols to be stored and 'paged' through to provide access to a larger virtual vocabulary.

Such systems undoubtedly offer improvements over previous incarnations. However, in general they are 'need led' as opposed to being research based, e.g. the need to increase the number of available symbols and hence vocabulary size; the need for *speech* output etc... As a result many research based questions remain unanswered, and many more are posed by adding layers of complexity to such originally elegant designs.

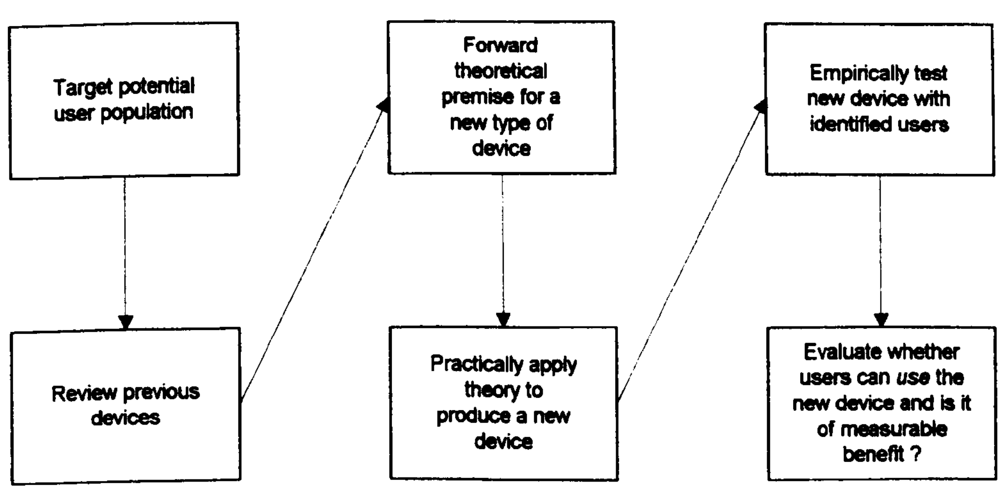
The objectives

The objectives of this thesis are manifold. The main objective however is to produce a software based ACD which is built around psychological theory and practical knowledge of the LD population who might benefit from such a device. That is, to design and implement an ACD that is tailored to it's users, rather than simply being a traditional approach implemented on modern technology. The overriding ethos is to provide the most benefit to users by practically applying psychological theory to push available technologies to their limits.

The envisaged system will be based on photorealistic digitised images as symbols, make use of digitised speech output, will have an in-built Automated Response Tracking System (ARTS) to keep track of user interaction and will have an interface based on a hypermedia metaphor.

The objectives of the thesis can best be summarised by the following sketch diagram :-

Fig 1.1, Sketch outline of objectives



The ultimate goal being to produce a software based ACD which is based in psychological theory and of tangible benefit to users. The resulting software package has become known as "Easy Speaker".

> Although many design and implementation issues have been implicit for a number of years it is only recently that they have been made explicit and more formulaic. During a recent research priorities workshop 14 leading American proponents of the AAC field formulated no less than 6 research priorities (Beukelman and Ansel 1995) :-
>
> ❶ To study the impact of AAC technologies on the development of communication, language, natural speech, and discourse skills of people with severe communication disorders.
>
> ❷ To study the influence of user variables (for example, knowledge, skill and learning style) on AAC system use.
>
> ❸ To investigate the impact of AAC system features on communicative competence and interactional skills of users.
>
> ❹ To develop tools and strategies to validly and reliably measure communicative, operational, linguistic, strategic, and social competence and of children and adults who use AAC systems.
>
> ❺ To investigate the effectiveness of AAC interventions by studying users of a variety of age, aetiologies, and social contexts and to determine those factors that are related to success and failure of AAC use.
>
> ❻ To encourage the academic development of researchers with a focus in AAC by establishing predoctoral and postdoctoral research and training opportunities.

Although the authors current research began informally some six years ago and formally in 1992 many of Beukelman et al's tenet's were shared implicitly and so in all probability evolved from common premises. With such commonality it is useful to use Beukelman et al's *accepted* priorities as a framework on which to expand the ideals and goals behind the development and application of Easy Speaker. Within the scope of a thesis one cannot hope to fulfil all sub-areas of such broad priorities, but instead to chip away at them.

❶ "To study the impact of AAC technologies on the development of communication, language, natural speech, and discourse skills of people with severe communication disorders."

Within this priority the need to implement the *correct* symbol set for potential users is crucial. In practice there are numerous symbol sets in everyday use around the world, ranging from the more abstract to the concrete; with many being idiosyncratic to the particular user. Examples might include Bliss, Rebus and ad-hoc hand drawn or pictographic symbology. However, such symbology can prove too complex for potential LD users so that they either give up, or only learn a rudimentary portion set needed for basic interaction. In general such systems tend to be learnt by rote and place a great deal of demand on the part of the user. Such demands may not be intellectually met by many. With regard to symbol set usage the author proposes from both theoretical and practical standpoints that an idealised symbol set should be constructed from photorealistic images of real objects, events, and people. Furthermore that these should be of around passport photo size and quality, in full colour, and should be easily digitised from the original photograph by use of a image scanner. As digitised images are used the user can actually be incorporated within the symbol set to increase salience. A concrete example might be a image of the user pictured drinking from a real can of Coke. When the symbol might be selected a piece of digitised speech might *say* "Can I have a drink of Coke ? - please". It is hoped that a base symbol set might be formed, which is then added to by the user or their immediate carers to build upon it. The aim is to provide a synergistic symbol set, that has a good learning curve, and incurs a low cognitive load in use. To use the old adage, "a picture paints a thousand words". Why put a LD user through the frustration and effort of learning that an abstract symbol represents a can of Coke, when you can have a photorealistic picture of a real can of Coke representing the real object ? Modern technologies have made such ideals feasible for a number of years, although now such capabilities are entering public domain.

Output from such devices is also considered as being crucial. Traditionally *voice* output has been in the form of synthesised speech which is generated phonetically from textual input. Although the latest synthesisers such as "DecTalk" are relatively advanced, listening to such output is still highly demanding for the majority of normal language users. However the widely forgotten flip-side is that ACD users themselves have to listen to any output too. As Beukelman et al acknowledge many users use a combination of natural speech with their AAC techniques. Recent research has suggested that synthesised speech is more cognitively demanding when listening to and understanding it. The author suggests that systems, such as Easy Speaker, should where possible use digitised human speech, which can by virtue of the medium carry all of the information contained in natural spontaneous speech. The argument being that it is *better* for listeners and for user alike.

The notion of ACD's acting as constant, almost covert, speech therapists is considered in-line with Beukelman et al's mixed mode assertion. The level of cognitive overhead imposed by synthesised speech is examined experimentally. The argument of having to "tune-in" to synthesised speech is also considered.

Some advanced commercially available ACD's can make use of both synthesised and digitised speech either exclusively or simultaneously. The notion of such mixed modes is also addressed. If digitised speech is utilised the syntactic details of constructing utterances is considered. For example, should utterances consist of whole pre-formed messages, be constructed from syntactic segments, or be made up from individual words ? For example :-

Whole	Can I have a drink of Coke ? Please.
Syntactic segments	Can I have - a - drink of Coke ? - Please.
Individual words	Can - I - have - a - drink - of - Coke - Please.

Thus the question of speed versus flexibility combined with increased complexity is raised, as is the question of syntactic learning.

Beukelman et al's third premise in this research priority is "system organisation". In most cases modern computer based ACD's are based on "dynamic displays" whereby symbol density is virtually increased by paging through various screens filled with symbols. Symbols can be selected as screens are paged through, thus increasing the vocabulary size available to the user. However many systems are simply based upon a book metaphor where pages are literally turned forward or backward to reveal more or less symbols which can then be selected. Although these pages might be organised in to meaningful wholes, with related symbols on each page, they are not coherently linked. Easy Speaker utilises a hypermedia style metaphor in system organisation. Whereby a user can traverse a large virtual three dimensional vocabulary of symbols by selecting special symbols which act as links to other related pages of symbols and links. One might view such a large intricately linked system as being a type of system "grammar" which the user must learn in order to use the system. The links between pages would themselves be digitised photorealistic images of exemplars of a given category. That is, they would be the archetypal member of a given category, or screen full of symbols. For example a photorealistic image of the user having breakfast surrounded with a red box might indicate that when selected the screen would change to symbols to do with "breakfast", e.g. breakfast cereals, drinks etc... when any of these symbols would be selected the digitised speech associated with them might be appended to the utterance under construction ready to be output when requested. In addition on any page, any mixture of links and speech symbols might appear. On the "breakfast" page for example there might be link symbols such as "lunch" and "dinner" etc... These might enable the user to select these so that he might be transported to pages where he might talk about what might be included in a later meal.

Again as with the symbol type themselves the author will forward theoretical notions of why one might wish to organise a large vocabulary based on such a methodology. For example, in relation to memory the system might be organised to represent the user own semantic memory, or knowledge hierarchy so that it might be traversed rapidly and for little cognitive effort due to it's synergistic nature. In addition the question of whether additions to the hierarchy would expand the users own knowledge and linguistic system are discussed. That is, does the system help form and strengthen the relationship between concepts in addition to providing a low demand access strategy ? Support for such a notion is taken both from theoretical works and from the authors own experimental work with real users of Easy Speaker. Experimental evidence is taken from the fact of whether or not users can learn to navigate a relatively complex hierarchy in order to construct set utterances. The more rapidly users' learn to use the system the lower the demand imposed in use would be presumed to be.

In general LD users have been precluded from using many AAC strategies and ACD's by virtue of their inherent complexity. By addressing these three key areas from both theoretical and research based standpoints one would hope to produce an ACD that on the surface appeared complex and offered greatly enhanced communication opportunities; but was so synergistic and of such low cognitive demand that all but the most intellectually and cognitive impaired could not make use of it. In sum the key areas which have been addressed within the first priority are :-

❶ Symbol type
❷ Output format
❸ System organisation

Indeed these might be seen as the three key tenets of any prospective ACD. All three must be as appropriate as is possible for the device in question to be of benefit.

❷ To study the influence of user variables (for example, knowledge, skill and learning style) on AAC system use.

With the realisation that it is near impossible to develop a system that is 'all things, to all men' designers are now beginning to realise that user variables are at least as important as system variables. To use an analogy, if the cockpit of a well designed Far Eastern plane is physically too small for Europeans, no matter how ergonomic the instrumentation they won't be able to fly it comfortably. Similarly an overly complex ACD will only ever be used effectively by the most able.

During the development of Easy Speaker the author has first hand experience that within such a user group it may not just be intellectual or cognitive variables that can affect ACD usage. Other factors such as physical constraints, ranging from the extreme of paralysis to the hindrance of motor tremor. Invisible physical impairments are also common, with uncorrected visual impairments topping the list. Where possible these have been catered for within the system.

Beukelman et al comment how, "Successful AAC use depends on a functional compatibility between the motor, cognitive, linguistic and sensory capabilities of the user and the operational requirements of the AAC system." By appropriate system and interface design the demands in terms of "operational requirements" can be reduced. However user variables are less predictable and their impact can only be assessed, slowly improved and consolidated with system usage. A strength in one area can compensate for a weakness in another to a certain extent. Throughout, the underlying theme that individuals are, by their nature, idiosyncratic has been maintained.

A major area of concern is whether motor learning, through the development of co-ordinative structures, develop within such individuals. Do motor skills hone with practice so as to approach the automaticity one might expect within an unimpaired individual ? And does learning follow a similar pattern ? Such questions are considered separately from any input device chosen. Devices might range from touch screens, to joysticks to mice. The focus is on strategy employed.

User competence, or skill, was formally assessed within Easy Speaker usage by use of the built-in Automated Response Tracking System (ARTS) which monitored interactions whilst users' completed a number of set tasks known as the Easy Speaker Assessment Tasks (ESAT). Such an approach enabled the author to follow an plot measures such as :-

- ❶ Cursor movement (distance cursor moved in mickeys, ≈ 1 mickey = 1 mm on screen)
- ❷ Highly accurate split timings (milliseconds), i.e. timings for each symbol selection
- ❸ Constructional latency for each utterance

By very accurately measuring the cursor track motor learning can be monitored. Cursor track, or tack, is used as a measure of efficiency, the less the cursor is moved to complete a set task the more efficient the strategy the user is employing. If levels of efficiency improve over time one can infer that motor learning has occurred. Also one can infer that higher levels of cognitive function have been called upon as planned strategies must be acted upon to increase efficiency.

Related to this is the use of highly accurate split timings. By examining split timings one can assess whether the user is also becoming faster, again implying some degree of learning and application of planned strategies. However one must bear in mind that speed might not be correlated with efficiency. A user may be very quick, but highly inefficient. Conversely a user may be slow, but demonstrate use of a highly efficient movement and selection strategy.

Such measures are generally more meaningful when the complexity of the task is varied. For example, the number of symbols that must be selected, and the number of virtual pages that must be visited to select the required combination of symbols. The task can then be assigned a numeric value indicative of its complexity. By using such measures one can estimate the level of "operational competence" (Beukelman et al 1995) a user has in relation to the system.

Crucial to the success of any ACD is the ability for it to reflect changes in the abilities of the user themselves. For example, as the user develops the vocabulary may need to be modified and/or expanded to suit. The process of vocabulary development might be viewed as being a never ending looped process where addition to the system might be driven both by the caregiver adding more complex symbols and links which might in-turn prompt the user to request more symbols and links, and vice-versa. For Beukelman et al "fast mapping" of a users "culminating experience and language base" is seen as essential. The system should grow to reflect the development of the user, and the ease with which such changes can be implemented is paramount.

Not only is the users *viewpoint* of the system important, but so too is the caregivers. Many systems employ complex scripting languages to program symbols and virtual screens; some even resort to the need to learn phonetic scripts. If any base system is too difficult for caregivers to modify, irrespective of user benefits, it is unlikely that its full potential will be realised. Such a view is vividly brought to life if one imagines a time/resource pressured environment such as the NHS where many devices might be used. Within a service setting what is needed is a base system that can easily and quickly be modified to suit a particular user. Many existing systems cannot.

As the "caregiver interface" is linked to any devices success, in relation to Easy Speaker, the views regarding the *programming* interface and *programming* method were assessed using the University of Maryland Questionnaire for User Interface Satisfaction 5.0 long (QUIS 5.0L). This was administered after initial usage and introductory *programming* of the system and again after around a weeks prolonged experience. This was done to evaluate potential caregivers 'first impressions', and then to see how these changed with a little more experience. Initial impressions could then be evaluated against later more experienced viewpoints to see if initial views were upheld or whether caregivers grew more knowledgeable of the system and found it more or less easy to utilise. Also areas or strengths and weaknesses would be signalled for future development.

As highlighted many in service professions may only have a limited time budget when learning to use any ACD, therefore a well designed *programming* interface, which can be rapidly learnt and successfully manipulated is vital. Just as success is motivating for users, so too is success for caregivers. In sum the key areas which have been addressed within the second priority are :-

- ❶ The influence of user variables of ACD usage
- ❷ Motor learning and skill acquisition
- ❸ Caregiver ease of *programming* (naive and experienced)

❸ To investigate the impact of AAC system features on communicative competence and interactional skills of users.

Again one of the burning questions of this priority is how to, "[give users] control over the language that is available to them" as Beukelman et al note. This is a question of interface methodology married with appropriate selection of input device. As noted earlier the objective is to give low load access to a potentially huge symbol based vocabulary. To achieve easy access the dynamic display approach that Easy Speaker utilises is based around a hypermedia metaphor; where intricate links can be formed between symbols and other screens. However many other devices use a simple paging arrangement, based on turning the pages of a real book, to reveal more or less symbols. Theoretically an approach such as Easy Speaker's should enable low load, almost synergistic, access. This should allow for greater communicative competence due to its similarity with many models of knowledge organisation. The organisation of symbols can theoretically be matched to the users current level of understanding allowing the next symbol in sequence location to be almost 'obvious' to the user.

In terms of input device Easy Speaker follows an open systems approach, whereby due to the Microsoft Windows operating system it runs under, any device that replaces a standard mouse in that environment can be used as a control device. As MS Windows is the best selling operating system world-wide this gives a potentially huge variety of "plug 'n' play" solutions that can be used. Generally these plug straight in to the standard mouse port without modification. These range from touch screens, pressure sensitive pads, virtual reality helmet style visors, and eye tracking spectacles to less space age devices such as cordless joysticks, radio controlled mice, and button panels etc... Easy Speaker also caters for keyboard input by means of the cursor keys, enter key and space bar. Although input devices are not tested empirically they are considered briefly in terms of the range available and their relative merits with respect to potential use population.

As most ACD's tend to use synthesised speech this is a major area of concern within this priority. Beukelman et al consider, "the influence of the rate of synthesised speech on listener comprehension and listener acceptability ?" With this question in mind an empirical study in to the cognitive load incurred when obeying synthesised and digitised task instructions was carried out. Although not strictly testing listener acceptability it does tackle the crux of the matter, listener comprehension. As noted earlier this is crucial if any device is to act as both ACD and covert speech therapist. By testing cognitive load one can propose whether or not synthesised or digitised speech is best employed for output within any ACD, including Easy Speaker. Together with other theoretical and research based findings a case is made against the *de facto* use of synthesised speech, especially for the user group in question.

Attention is also drawn to the ease with which users can learn to use the system and it's symbology. As regards system learning this is done by empirically examining users' performance on a number of set tasks and noting performance gains. Various symbol systems are evaluated on the basis of previous theoretical and research based work for such a user group. This is taken as the basis for suggesting that a photorealistic set of symbols be utilised whenever possible.

❹ To develop tools and strategies to validly and reliably measure communicative, operational, linguistic, strategic, and social competence and of children and adults who use AAC systems.

Ongoing assessment is crucial to the success of any ACD. This enables modifications to the system software to be considered and affects of such modifications evaluated. It also allows for caregivers to focus on areas of weakness and to consolidate strengths. Improvements can be plotted and used to give positive feedback to the user.

Beukelman et al consider "operational competence" and the tracking of operational competence as being invaluable. Such a notion of on-line tracking is considered essential. In Easy Speaker all interactions with the system are recorded in the minutest detail ready for inspection at a later date by the Automated Response Tracking System (ARTS). These include what is output, when it is output, how long it took to generate and measures relating to the effort put in to generate that output, e.g. the volume of cursor movement. Fundamental descriptive statistics are also calculated, e.g. how many seconds were needed to produce each word of output etc... In addition useful predictions are made, e.g. the estimated number of words per minute based on the last utterance generated. Using such an integrated tracking approach allows for set tasks to be given which can be used along with ARTS to assess the users level of current "operational competence". Such tasks can also be used to evaluate users' knowledge of the hierarchy itself after a period of acclimatisation.

However such measures are mainly concerned with performance, learning and knowledge of the system. These are viewed as primary goals by the author. Secondary goals on the other hand relate to off-device improvements in language, motivation, cognition and sociability. Secondary gains were evaluated by means of measuring functional levels of behaviour in these areas quasi-experimentally. This was done by using 67 items taken from the Adaptive Behaviour Scales (ABS) and the Functional Performance Record (FPR). The modified questionnaire was given to immediate care-givers of each user prior to intervention and again after 6 months of Easy Speaker usage. Each item was then examined to check for improvement, worsening or no change.

Thus two measures of Easy Speaker usage were examined. Firstly pure performance measures of system competence, and secondly off-device transferable benefits in language, motivation, cognition, and sociability.

❺ To investigate the effectiveness of AAC interventions by studying users of a variety of ages, aetiologies, and social contexts and to determine those factors that are related to success and failure of AAC use.

The effectiveness of any AAC/AAD approach is *only* lent credence when *trialed* with a reasonable number of real users who differ in aetiology, abilities and needs. In addition social factors provide for other potential additional/confounding variables. In this respect Beukelman et al note how "In the current climate of accountability in both education and medical care arenas, the need for efficacy or outcome research is obvious." Although suggestive of a, "bang for the buck" mentality, the assertion of close monitoring of outcome is seen as imperative. To date, perhaps due to the 'newness', of the area there are *no* long term studies of ACD effectiveness and impact on users' everyday lives. By long term one might consider at least two to three years and upwards of experience with the same device.

Also considered within this priority was the need to consider the, "impact of AAC on family structure and dynamics in terms of levels of stress, changes in family goals for the user, and changes in user preferences.", Beukelman et al (1995).

Such long term goals are beyond the scope of this thesis which concentrates on design, implementation and monitoring issues associated with modern ACD's. However with the increasing uptake of such devices such long term research must be carried out once design and implementation issues have been resolved.

❻ To encourage the academic development of researchers with a focus in AAC by establishing predoctoral and postdoctoral research and training opportunities.

Beukelman et al in common with the author notes how the field of AAC is still relatively young and that the research community built around it remains quite small. Also of note is that the view that, "the rapid growth of the field was accomplished primarily by the transition of clinicians, engineers, educators, and administrators with little or no specific AAC training into this new developing area."

Such views can be seen in many feature rich, as opposed to functional theory based and field trialed ACD's. As Newell and Alm (1994) point out, "[often] designers have assumed that their users will have the characteristic of a 25 year-old male who has a PhD in computer science and is obsessed with technological gadgets rather than getting on with the job. Systems are a triumph of functionality over usability." Although this quote is taken slightly out of context its meaning is not, with many ACD still remaining usable only by the most able. This is especially true of the LD population who might benefit most from a well designed ACD.

The goal of this final priority is to move away from generic approaches to design and to tailor devices toward the intended user populations. This should be done through both theoretical ideals and research based investigations. Devices should be field trialed throughout their development and users' suggestions should be taken on-board. Perhaps the best example of this is provided by McGregor and Alm 1992, who's own research team included a member with a communication dysfunction.

With raised interest in the field more in-depth research will be conducted and recognised by many of the commercial developers who continue to produce inadequate ACDs for the majority of the LD language impaired population. The ethos being to produce good devices and AAC strategies by following academic theory and research principles to a commercial conclusion. Hopefully as such a precedent is taken on board more and more researchers will become interested and involved in the field.

SUMMARY

Hopefully this thesis will hint at the direction future research and commercial implementation might follow. The specific objectives of this thesis are summarised below :-

❶ To examine the handicaps and abilities of the potential user population

❷ To review previous commercial ACD's, generically then specifically

❸ To suggest and implement a symbol set based upon psychological theory and practicability

❹ To suggest which access strategy, or interface metaphor is most synergistic with users abilities

❺ Propose which type of output for an ACD is generally most suitable for listeners and users alike from a cognitive load prospective, i.e. synthesised or digitised speech

❻ To design a device which takes account of goals 1-5 above

❼ To empirically test whether real users can learn to use such a device and make constructive use of it

❽ To objectively measure any off-device tangible benefits in language and cognition, e.g. language, memory, motivation etc... as a result of medium-term usage of such a device

❾ To extrapolate findings to suggest future research and commercial directions for 'new wave' ACD's

CHAPTER 2: Characteristics of potential ACD user populations

If any ACD is to succeed in its role it must be designed with the abilities and limitations of the potential user population in mind. Although within any population there is likely to be large variations in needs and abilities it is useful to design for the *norm*.

In the case of Easy Speaker its intended user population are those with a Learning Disability combined with expressive language and vocal impairments which makes normal communication difficult, slow, or laboured. Intellectual or physical impairments may mean that a standard signing system is unable to be utilised, e.g. Mackaton etc... Most are assumed not to have sufficient grasp of English to be able to read or write, and currently don't use any alternative method of communication other than telegraphic spoken English.

Despite many individuals fitting this profile what is needed is a more detailed description of general abilities and limitations, as well as common aetiologies and secondary conditions. The relevance of ACD intervention is also be stressed in relation to these.

Although within this chapter there will be the expected quotations of facts and figures regarding the abilities of this group, the focus will be on abilities and limitations and not epidemiological data per se. It is pointless to quote statistics on measures such as IQ, Mental Age, detailed language profiles and for that matter aetiology. What is important however is what individuals are capable of and what their limitations might be.

INTRODUCTION

A key tenet in designing anything new is to know who prospective users are likely to be. In the case of an ACD this is especially pertinent. One must not only know how many potential users there might be, but in the case of Learning disabled users how they are stratified in terms of ability. One must also know what their current levels of communicative effectiveness is likely to be. Crucially linguistic and cognitive handicaps must be understood for a device to be designed to suit existing levels of ability. Above all the relevance of ACD intervention must be taken account of. Thus the 'background' of potential users must be understood.

BACKGROUND

Prevalence of communication disorders within the Learning Disabled population

It has been accepted for many years that many Learning disabled (LD) individuals are likely to suffer from significant communication deficits (e.g. Bateman 1964, Marge 1972). These can be either as a result of their intellectual and cognitive impairments, physical disabilities, or a combination of the two.

However up until relatively recently evidence for such assumptions within the LD population was descriptive, anecdotal, or derived from very small samples (Feagans 1983). One of the first studies to attempt to measure prevalence within a reasonably sized population is that of Gibbs and Cooper (1989). Gibbs and Cooper measured, using standard tests, articulation, fluency, voice, language and hearing disorders in a population of 242 LD children aged between eight and twelve. A summary of Gibbs and Coopers results are shown below :-

Fig 2.1, Prevalence of communication disorders in a typical LD population

Disorder	No of subjects	Prevalence (%)
Articulation	56	23.1
Language	219	90.5
Fluency	3	1.2
Voice	29	12
Hearing Puretones	18	7.4
Middle Ear Function	38	15.7

From Gibbs and Cooper (1989)

The startling finding of their investigation was that 96.2% (233) of the 242 LD individuals studied had one or more communicative disorders. No age or sex related differences were significant. Despite such a high percentage, only 6% (14) were receiving any speech-language pathology services. In the main those that were concentrated on articulatory problems.

Gibbs and Cooper conclude that with such a high proportion of individuals having some degree of communicative disorder that, "Learning Disability was practically synonymous with a diagnosis of a communication disorder." The potentially most worrying aspect of their findings is that the prevalence of speech and language disorders did not decrease as function of age. Within normal populations the frequency of speech and language problems decreases as the population grows older. This is suggestive that these prevalence rates can be extrapolated to older populations of LD individuals.

The prevalence of articulation disorders, 23.1% (56), is perhaps indicative of the size of the LD population that might benefit from an interventional strategy such as Easy Speaker. An educated guess might suggest that the most severely affected account for around 5% of the whole LD population. Obviously this group would be likely to benefit most for a low demand ACD like Easy Speaker.

Communicative effectiveness

Having established the estimated frequency of language disorders it is useful to examine how effectively, despite their handicaps, such individuals can, "get the message across" (Blackwell, Hurlburt, Bell et al 1989). Blackwell et al were primarily interested in the effectiveness of communication within a LD population. They defined an effective communicator as, "one who can initiate a message and respond to a message using a multi-modal approach of verbal and non-verbal behaviours". The goal of their research was to :-

- To obtain more detailed information on the range and extent of communication difficulties in people with a mental handicap
- To pinpoint more accurately what constitutes a communication difficulty
- To identify factors contributing to such difficulties

Blackwell et all studied some 563 adult individuals who were either in hospital settings or who attended day centres were assessed by giving staff a two-part questionnaire. The first half related to each individuals ability to :-

- Express basic needs
- Use communication in social interaction

On the basis of scores on this first questionnaire individuals were assigned to groups 0, 1 and 2. The lowest third, mid-third, and top-third respectively. From each of these 50 individuals were then selected at random.

On the basis of these 150 individuals Blackwell et al constructed the following table :-

Fig 2.2, Stage 2 sample, group, location and sex

		Hospital	Day Centres	Totals
Group 0				
	Males	15	9	24
	Females	20	6	26
Group 1				
	Males	16	14	30
	Females	15	5	20
Group 2				
	Males	18	5	23
	Females	9	18	27
	Totals	93	57	150

The second half of the questionnaire was then given to the immediate carers of this group of 150.

The questionnaire assessed each individuals ability in the following areas :-

- A developmental measure designed to identify whether individuals were functioning at a level of symbolic understanding and thus able to appreciate "word" labels.
- A comprehension section.
- Early situational understanding of "everyday" requests
- "Derbyshire Rapid Screening Test" to assess verbal language comprehension
- An imitation section
- Imitation of generalised hand movements such as clapping to specific sign imitation
- Imitation of orofacial movements
- Imitation of individual speech sounds
- A section considering communication behaviours of the individuals
- Spontaneous imitation versus prompting and necessary repetition in order to engage the individual in discourse
- The functioning of communication which the individual used, ranging from "babble/talking to self" to "sharing experiences"
- The manner in which these intentions are communicated, e.g. ranging from eye gaze to formal sign or connected speech
- Whether verbal communication had :-
- Developing syntax
- Impaired phonology
- Impaired fluency
- A measure covering four main areas of effectiveness in the environment - to assess, opportunities available to the individual, and how successful the individual was in capitalising on such opportunities
- Freedom of movement
- Freedom of choice
- Handling money
- Responsibilities in the home
- A questionnaire covering the individuals medical history

As can be gleaned from list of goals the questionnaire was geared toward the assessment of adaptive, or functional behaviour as a measure of ability. Such an approach is common within the LD field with individuals being unlikely to be successfully assessed using more formal techniques. A similar approach was taken by the author in chapter 10, when assessing Easy Speakers users off-device gains in language and cognition.

Blackwell et al derived six important finding from the questionnaires. The first being that their was no significant sex difference in communicative ability. However they did suggest that women are more likely to make use of any basic skills necessary for social interaction. This finding tallies with that of Gibbs and Cooper (1989).

Despite the sample ranging from 19 to 89 years of age, with a mean of 37, there was no significant age difference in communicative ability. Such a result was hinted at by Gibbs and Cooper when they suggested that speech and language disorders were unlikely to decrease as a function of age within a LD population.

One might also assume that medical condition, or aetiology might affect the individuals communicative ability. However, Blackwell et al found that it did not. The scoring out of a maximum 134, was distributed as follows :-

Fig 2.3, The effect of medical condition of questionnaire score (out of possible 134)

Medical condition	Median questionnaire score
No medical condition	64.5
Epilepsy	50
Down's syndrome	81
Other	58

Although medical condition seems to be unrelated to communicative ability, Blackwell et al found that physical handicap was significantly related. Individuals with a physical handicap scored significantly lower, in terms of communicative ability than those with no handicap. Blackwell et al nificantly lower, in terms of communicative ability than those with no disability. Blackwell et al point out however that the equation, "physical handicap = poor communication ability" grossly over-simplifies the point. Blackwell cite McGarry and West 1975 who suggested that interactions between normal people and those with some physical disability are characterised by reduced spontaneity and guarded stereotypic behaviours. Such interactions are obviously not the best foundation for developing communication skills. Thus rather that being a direct function of physical handicap, it is perhaps a combined affect of previous impoverished communication opportunities and lack of social learning.

As one might expect whether individuals were verbal or non-verbal had a statistically significant effect on communicative ability. That is, when one takes into account not just verbal aspects of communicative interactions per se. Within this area only groups 0 and 1 were compared, as group 2, the more able individuals, were all verbal. Blackwell et al broke down verbal ability by group to produce the following table :-

Fig 2.4, Number of verbal and non-verbal individuals by group

	Group 0	Group 1	Group 2
Verbal	18	38	50
Non-verbal	32	12	0
Totals	50	50	50

As can been seen from the table above verbal ability is highly related to communicative ability.

Location also proved to be a significant factor in communicative ability, i.e. whether individuals were located in day centres or hospitals. From the median scores obtained Blackwell et al constructed the following table :-

Fig 2.5, Median scores for individuals based in day centres or hospitals by group and location (out of possible 134)

Group	Day centres	Hospital
0	70.5	16.5
1	80.5	47
2	106	84

Those individuals who visited day centres communicative ability was significantly above those of similar disposition who were hospitalised. However, complicating the effect of location is the fact that 86% of the non-verbal population were located within a hospital setting. This may suggest that those with the severest communication difficulties were more likely to be hospitalised than the vocal. Alternatively due to the being placed in a *uncommunicative environment*, the result might be few communication opportunities, which in turn leads to poor communication skills development. Hence their initial difficulties might be exaggerated. In reality an interaction of the two is more likely. Interestingly hospitalised population shared a mean age of 41 years whilst those attending day centres shared a mean age of 31 years. Again this is suggestive that those with more marked communication difficulties tend to 'gravitate' toward hospital settings with increasing age.

Further evidence that hospitalised individuals tend to be more likely to be communicatively disadvantaged comes from the finding that their environmental effectiveness scores were significantly below that of those who attended day centres.

Fig 2.6, Median scores of environmental effectiveness by group and location

Group	Day centres	Hospital
0	23	2
1	25	8
2	40	22

This is suggestive that those who are most communicatively disadvantaged are also poor manipulators of their environment. For example, they exert little choice, have poor freedom of movement, and tend have few responsibilities. Again within a hospital setting these tend to be more severe.

Interestingly Blackwell et al note how the researchers ratings of communicative behaviours were significantly different from those of the immediate caregiver(s) who rated each individual on the questionnaires. These differences were greatest for the less able group 0, and decreased as communicative ability improved in group 1, and virtually disappeared when rating group 2, the most able. Blackwell et al suggest that within less able groups caregivers interpret behaviours which are not clearly expressed as communicative. The result is shared meanings, and patterns of communication which are almost 'secret' in nature to those outside that immediate circle.

Blackwell et al's conclusions are summarised below :-

- Those involved with people who have a mental handicap should adopt a generous definition of communication and recognise that effective communication is not the sole responsibility of the individual a mental handicap
- The results empathise the interactive nature of all the skills involved in communicative ability. Thus, improvements in language usage are more likely to stem from programmes working on the entire process of communication
- Medical condition and physical disability do not necessarily hinder communication ability. Physical disability may have a detrimental effect on communication ability. Particularly when the disabled individual is also non-verbal
- The ability to use language has a greater effect on communication ability than physical disability
- Individuals at day centres showed significantly greater communication ability than those in hospital
- At all levels of ability, individuals in hospital had significantly less environmental opportunity than those at the day centres
- Increased environmental opportunity is correlated with improved communication ability
- Being verbal enables individuals to capitalise more fully on environmental opportunity
- Opportunities for environmental control (making choices and decisions) should be more readily available. Such opportunities result in the naturalistic settings for the development of communication ability

Common language and cognitive handicaps

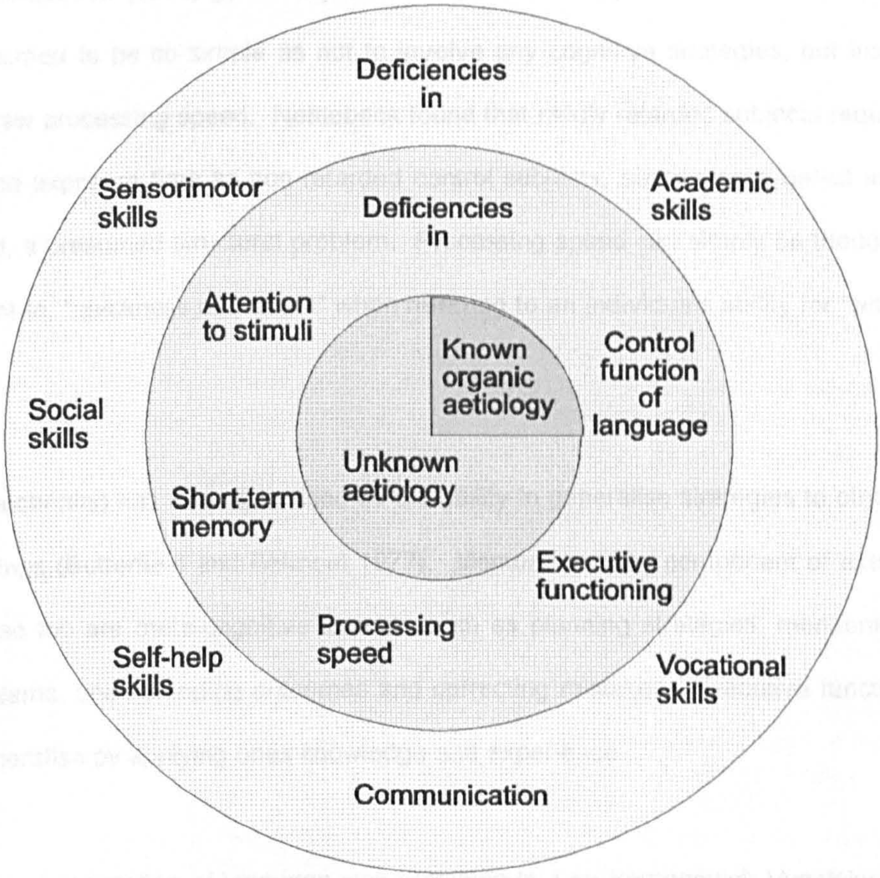
Other researchers in the field have produced profiles, or meta-analyses, that attempt to list the features that might characterise those with a LD and communicative disability. Westwood (1993) for example lists some eleven difficulties which are common to those with any degree of LD. These difficulties may be present in any combination and are of variable severity :-

- A history of late speech development (and continuing immaturities in articulation and syntax)
- Visual perception problems (frequent reversal of letters and numerals - distorted or blurred word shapes)
- Auditory perceptual problems (including difficulties in identifying sounds within words and blending sounds into words)
- Poor integration of sensory information (e.g. can't easily learn to associate and re-member printed symbols and their spoken equivalent)
- Weak lateralization (e.g., underdeveloped hand-eye preferences; directional sense confusion)
- Some signs of neurological dysfunction
- Hyperactivity
- Weak sequencing skills (as reflected in jumble letter sequences in spelling or in word attack skills in reading)
- Poor co-ordination
- Low level of motivation
- Secondary emotional problems due to learning failure and poor school progress

Given the obvious handicap many of these may impose it is easy to envisage how Gibbs and Cooper and Blackwell et al's subjects performed so poorly in terms of communication and related measures.

In a similar vain Davision and Neale (1990) proposed a visual scheme to illustrate the nature and aetiology of mental retardation.

Fig 2.7, The causes of mental retardation and the adaptive skills and general cognitive abilities that are wanting in retarded people



Adapted from Davison and Neale (1990)

Although their visual model shares many of the properties hinted at in earlier studies and to some extent made explicit by authors such as Westwood (1993), others are introduced. Namely, "Processing speed", "Executive functioning", and "Control function of language".

Processing speed is simply a measure of the speed at which an individual can process information. It is analogous to a computers raw processing speed, e.g. processor speed in Mhz. Again using the computer analogy, a computer can do things more quickly by one of two methods. By having highly optimised software algorithms; or by having a very fast processor. In combination each can compensate for the other. In human terms the software, or knowledge, is acquired and refined by learning, and the hardware is, neuroanatomical, or structural.

In measuring raw processing speed the goal is to keep cognitive strategies (control processes) to a minimum. Nettlebeck (1985) gave subjects a simple timed discrimination task (inspection time) that was assumed to be so simple as not to involve any cognitive strategies, but instead was a measure of raw processing speed. Nettleback found that mildly retarded subjects required around twice as much exposure time as non-retarded control subjects, suggesting a deficit in their processing speed, a presumed structural problem. Processing speed can simply be thought of as being equivalent to, "quickness of thought" when referring to an individuals ability for "witty repartee" for example.

Executive functioning has been described as the ability to generalise strategies to other times and to other settings (Butterfield and Belmont 1977). Memory is a key component of executive functioning, but so too are meta-cognitive aspects such as planning strategies, monitoring progress, solving problems, understanding outcomes and correcting mistakes. Executive functioning is the ability to generalise by applying ones knowledge and experience.

Finally, the control function of language was proposed by Lev Semenovich Vygotsky (1896-1934) who proposed that language and thought were initially separate entities but through development they became inextricably linked. He theorised that private speech separated from the public and assumes a control function. The control function of speech is considered as being crucial to intellectual behaviours and development (Vygotsky 1978). It is suggested that within the LD individual "inner speech" is never fully achieved and so limits ability to varying degrees.

A review of the Learning Disabled population would not be complete without the reference to standard measures of intelligence. However within the LD field these are less commonly used in isolation, as compared with measures of adaptive behaviour. None-the-less they do serve as indicators and are useful for defining clear cut boundaries between sub-categories of severity, when combined with adaptive measures.

Classification and epidemiology

Perhaps the most important classificatory system within the psychiatric and clinical psychology field is the DSM-IIIR, which classifies psychiatric conditions through to defining Learning Disability strata. The DSM-IIIR defines four bands of severity based on IQ measures. It should be remembered that an IQ measure tells us nothing of an individuals *real* abilities. The four sub-divisions based on IQ are :-

Fig 2.8, DSM-IIIR classification of Learning Disability based on IQ

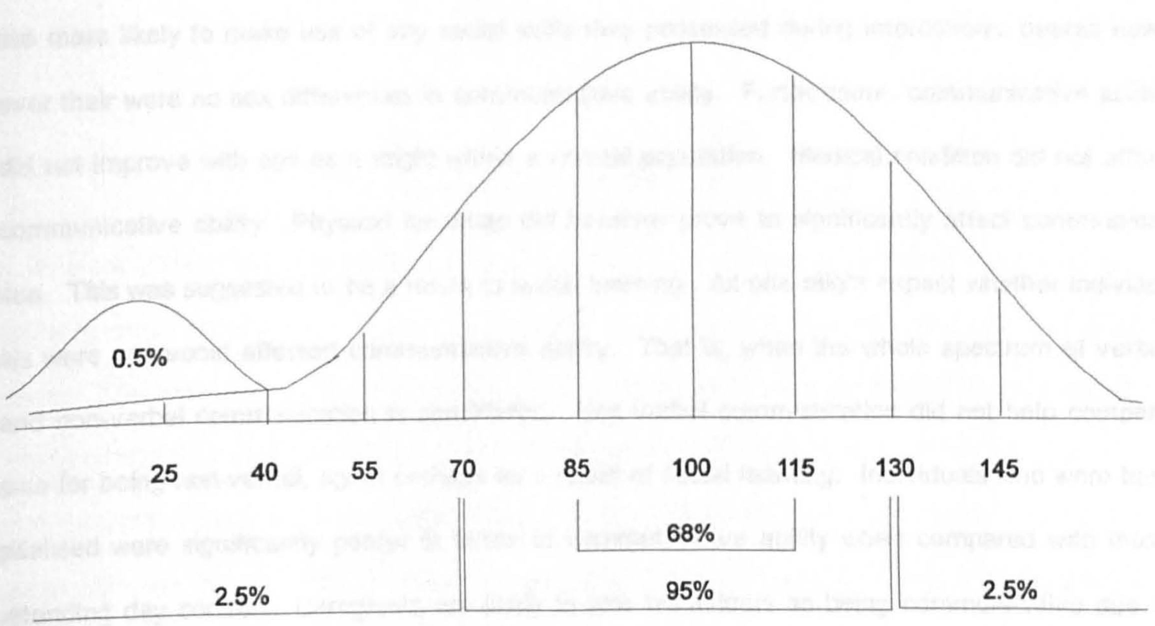
IQ and Label	Prevalence	Description
50-55 to 70 Mild Mental Retardation	85%	They are not always distinguishable from normal youngsters until they enter school. By their late teens they can usually learn academic skills at about sixth grade level. As adults they are likely to be able to maintain themselves in unskilled jobs or in sheltered workshops, although they may need help with social and financial problems. Further, they may marry and have children of their own. Only about one percent are ever institutionalised, usually in adolescence for behaviour problems. Most of the mildly retarded show no signs of brain pathology and are members of families whose intelligence and socio-economic levels are low.
35-40 to 50-55 Moderate Mental Retardation	10%	Brain damage and other pathologies are frequent. The moderately retarded may have physical defects and neurological dysfunction's that hinder fine motor skills, such as grasping and colouring within lines, and gross motor skills, such as running and climbing. During childhood these individuals are eligible for special classes in which the development of self-care skills rather than academic achievement is emphasised. The moderately retarded are unlikely to progress beyond second grade level in academic subjects and can manage this learning only later in childhood or as adults. They may, however, learn to travel alone in a familiar locality. Many are institutionalised. Although most can do useful work, few hold jobs except in sheltered workshops or family businesses. Most live dependently within the family or in supervised group homes. Few have friends of their own, but they may be left alone without supervision for several hours at a time. Their retardation is likely to be identified in childhood, for their sensorimotor coordination remains poor, they are relatively slow to develop verbal and social skills. In contrast to mildly retarded children, moderate retardates and those more seriously retarded are found in all socio-economic groups.
20-25 to 35-40 Severe Mental Retardation	3-4%	They commonly have congenital physical abnormalities and limited sersorimotor control. Genetic disorders and environmental insults, such as severe oxygen deprivation at birth, account for most of this degree of retardation. Most are institutionalised and require constant aid and supervision. For children in this group they to able to speak and take care of their own basic needs requires prolonged training; the self-care training that is provided within the special classes within the school system is usually inadequate except for the upper portion of this group. As adults the severely retarded may be friendly but can usually communicate only briefly on a very concrete level. they engage in very little independent activity and are often lethargic, for their severe brain damage leaves them relatively passive and the circumstances of their lives allow them little stimulation. They may be able to perform very simple work under close supervision.
Below 20-25 Profound Mental Retardation	1-2%	Generally require total supervision and often nursing care all their lives. Intensive training may improve motor development, self-care, and communication skills. Many have severe physical deformities as well as neurological damage and cannot get around on their own. There is a very high mortality rate during childhood.

Adapted from Davison and Neale (1990)

Despite being able to define the characteristics of the LD population, accurately forwarding a population estimate is near impossible. Most in the field, due to difficulties in measurement, accept a calculated population estimate based upon the normal distribution for IQ scores. The only more reliable figures are for those who are severely to profoundly handicapped resulting from brain damage at birth.

Based on the DSM-IIIR definition of mental retardation approximately 2.5% of the general population world-wide have a mild to profound degree of mental retardation due to having IQ's of 70 and below. That is, when calculated based on the theoretical distribution of IQ scores :-

Fig 2.9, The theoretical distribution of IQ scores



In the UK alone this would suggest a LD population consisting of some 1.45 million (2.5% of 57.5 million) when based on the expected distribution of IQ scores and using DSM-IIIR with a cut off of 70 IQ points and below. According to Gibbs and Cooper (1989) one might conservatively expect around 90% of those individuals to posses some degree of communication disorder, or well over a million individuals. With articulation disorders prevalence being around 23%, or around 325,000 individuals. The bump at the left hand side of the predicted distribution of IQ is caused by those individuals who as a result of brain dysfunction or injuries have severe or profound mental handicaps.

Summary of the abilities of those who might make use of an ACD such as Easy Speaker and the potential implications they might have

On the basis of research data one can conclude that the vast majority of the Learning Disabled population have one or more communicative disorders. Some authors go so far as to suggest that a Learning Disability is synonymous with that of a language disorder, Gibbs and Cooper (1989). Within the LD population it is suggested that around a quarter have marked articulatary problems, ranging from a stutter, to being non-vocal.

Other authors have chosen to examine "communicative effectiveness" and not just linguistic components of communication per se. Blackwell et al (1989) obtained detailed information on the range and extent of communication effectiveness within a LD population. They found that the less communicative an individual the more likely they were to be hospitalised. They found that women are more likely to make use of any social skills they possessed during interactions, overall however their were no sex differences in communicative ability. Furthermore, communicative ability did not improve with age as it might within a normal population. Medical condition did not affect communicative ability. Physical handicap did however prove to significantly affect communication. This was suggested to be a result of social learning. As one might expect whether individuals were non-vocal affected communicative ability. That is, when the whole spectrum of verbal and non-verbal communication is considered. Non-verbal communication did not help compensate for being non-verbal, again perhaps as a result of social learning. Individuals who were hospitalised were significantly poorer in terms of communicative ability when compared with those attending day centres. Caregivers are likely to rate behaviours as being communicative due to developed shared meanings, e.g. eye brow raising indicating a need etc... when *outsiders* may not.

In terms of common language and cognitive handicaps Westwood (1993) noted a history of late speech development; visual perception problems; auditory perceptual problems; poor integration of sensory information; weak lateralisation; signs of neurological dysfunction; hyperactivity; weak sequencing skills; poor co-ordination; low levels of motivation; and secondary emotional problems.

Davision and Neale (1990) noted deficits in academic skills; vocational skills; communication; self-help skills; social skills; sensorimotor skills; control function of language; executive functioning; processing speed; short-term memory; and attention to stimuli.

The DSM-IIIR in relation to learning disability classification was reviewed in order to give IQ ranges, prevalence rates and verbal description of the four strata of mental retardation. Namely mild, moderate, severe, and profound.

Using the normal distribution for IQ scores it was predicted that 2.5% of the normal population, on the basis of IQ, could be deemed as having a Learning Disability. Due to the lack of accurate figures for prevalence rates in the general population this percentage was extrapolated to calculate that approximately one and a half million individuals in the UK could be labelled as being Learning Disabled.

THE RELEVANCE OF ACD INTERVENTION

With some 23% of LD individuals having some degree of articulatary disorders, ranging from the slight to being non-verbal, the need for a VOCA (Voice Output Communication Aid), or ACD is clear. The need for some kind of ACD is reinforced when one considers the low levels of "communicative effectiveness" in relation to this group. That is, when verbal and non-verbal communication ability are considered. Authors such as Blackwell et al (1989) have suggested that within the least able the lack of communicative experience can actually hinder their progress and make them less likely to attempt to interact communicatively. For the least able non-verbal, a suitably designed ACD, offering even apparently simple discourse, may well improve their levels of communicative effectiveness. For example, they may be able to indicate simple needs or wants that may have previously gone unnoticed, e.g. requesting drinks when thirsty etc...

The more able non-verbal may be able to use such a device as a full-blown communication aid, used in everyday interactions. Research has suggested that the more communicative opportunities within a natural environmental setting the better. Beukelman et al (1995) stress how LD ACD users tend to utilise a mixed mode of ACD use, combining device interaction with use of whatever residual verbal and non-verbal ability an individual possesses. One might propose then, that a suitable ACD might both enable communication, but also trigger and promote natural residual communication abilities. Logically, prolonged use of such a device offering voice output may result in the users own expressive language ability being improved. That is, the device may act as an almost covert speech therapist. Thus, for those with less marked articulatary disorders such a device may be both of communicative and therapeutic benefit.

With such a high prevalence any device which could potentially improve language ability, whether it be used primarily as a communication medium or not, may be useful. However, one must remember that it is not simply communication skills that are likely to be affected by a LD. Motivation, attention, memory and cognition etc... are also typically impaired (Davison and Neale 1990 and Westwood 1993). These issues are of crucial importance when one considers any ACD's vocabulary representation and organisation, and general design issues. Such details are addressed within chapters four and five respectively. However it is useful to discuss briefly the *non-language* impairments that are typical within a potential user population, and the effects they might have on ACD utilisation.

One might expect that cognitive impairments may be more important than psychological variables such as motivation in ACD usage. However, many authors in the LD field acknowledge the low levels of general motivation and the common failure cycles displayed by many individuals. For devices, one to be used, and two to be effective they must be motivating. To impact on cognitive deficits they must enhance motivation, be rewarding to use, and act a self-reinforcers. Such reinforcement is crucial as many individuals display poor integration of sensory information. For example, they can't easily learn to associate and remember printed symbols and their spoken equivalent.

Related to motivation is the fact that many of those with a LD are prone to have a short attention span combined with proneness toward hyperactivity, or what at best might be described as overactivity. The devices motivational/reward component must be sufficiently high to offset reduced attention spans, and as a result lessen the tendency to become hyperactive. If the user can remain focused, concentration can be given to operating, and learning from the device in question. If the device is too difficult to use or unrewarding/unmotivating users focused attention is likely to broken. Overactivity and hence increased error rates may be the result. If this occurs any learning is unlikely to be reinforced, and thus remembered.

In order for motivational levels to be maintained users should find the device so easy to operate that they can't fail. Many authors advocate that LD individuals should move away from a "cycle of failure", where one failure leads to decreased motivation, and hence more failures and so on ... Failures should be able to be monitored and appropriate action taken to ensure that the user doesn't perpetually fail in what they want to achieve. Action might involve teaching the user a new strategy for completing the task, tailoring the device to the user, or redesigning the system. Increased device complexity obviously increases the likelihood of failure, so simplicity is the key.

Another common misconception is that LD individuals have *normal* auditory and visual perception. Research suggests that this is not the case with auditory and visual perceptual problems being common. Such problems may be either physical, e.g. ophthalmic, or neurological in nature. Obviously if these are remain uncorrected they must be encompassed with the design of an ACD. A standard "eye test" for a severely LD, non-verbal individual may not be practicable. Although use of an ACD cannot correct physical abnormalities, it can help sharpen and enhance residual levels of functioning to apparently raise levels of functioning.

Combined with poor perceptual skills many LD individuals also have poor co-ordination skills. These can either be as a result of neurological damage, or simply lack of environmental opportunity. Many individuals motor skills can be improved through extensive training, up to the limits of their potential. Some degree of co-ordination is essential for interaction with an ACD, and so devices should not require extremely, high accuracy, movement and target selection.

Weak sequencing skills are also common. This may make learning and utilising the "grammar" of any device difficult. Just as the individual may have difficulty in sequencing words together to make sentences, they may too in sequencing plans/strategies to operate the device in order to produce *spoken* output. This may result in slowed output rates, or at worst failed output. This means that the organisation, or grammar, needed to operate of any ACD should be as synergistic as possible with the users understanding and current levels of functioning.

Related to poor sequencing skills are general deficits in short term memory observed in many individuals with a LD. This typically means that any material which is overly long, too complex, or requires advanced manipulation is likely to be to overload the limited abilities of the individual. Sequencing skills are further affected if the sequence in question is overly long or complex. ACD's should therefore where possible not require users to remember and manipulate long sequences of actions that need to be taken to achieve a desired goal. If such sequences are long they perhaps might be guided, or prompted, by means of cues build in to a devices interface. Such cues might act as an aid to memory, and hence help in sequencing.

Some authors have suggested that "processing speed" (Nettlebeck 1985), or quickness of thought, within the LD population is slowed. The more severe the LD, the slower information processor the individual is assumed to be. In practical terms this may mean that complex actions needed to operate an ACD successfully may be slowed by virtue of each component action taking longer to 'think about', and therefore to complete. In computing terms the individual has, less computing power. Combined with other characteristics such as, short attention span, poor sequencing skills, poor co-ordination etc... this may decrease performance still further. To overcome any deficits in processing speed, more efficient strategies can be taught and used to operate the ACD. This is the equivalent of optimising a piece of computer code, or algorithm, in order to make it run more quickly without the need to increase the speed of the CPU.

Allied to processing speed is the common failure by LD individuals to display "executive functioning", Butterfield and Belmont (1977). Simply this is the inability to generalise strategies to other times and to other settings. In relation to processing speed, using simple efficient strategies for ACD operation can help compensate for a general slowness. However, to do this the individual must be able to generalise any strategies they are employing to utilise the device more productively. As memory is a key component strategies should be simple in order for them to be generalised, and used from occasion to occasion, and setting to setting. To promote use of such strategies the operation of the device should be as simple as possible.

Within the LD population language is inextricably linked with other cognitive and physical deficits. Whether or not language ability is a casual factor as Vygotsky (1978) suggested, a resultant deficit, or a parallel deficit is unimportant. What is, is the acknowledgement that language and other deficits seem to appear in combination frequently. With such apparent clustering some authors suggest a multisensory approach to language improvement, whether it be language based around ACD usage, the users own language, or a combination of the two.

Blackwell et al (1989) for example suggests that, "improvements in language usage are more likely to stem from programmes working on the entire process of communication". Such an assertion strengthens the argument that such a device should also, if possible, produce secondary benefits in motivation, attention, memory and cognition. Blackwell et al also suggested that increased environmental opportunity is correlated with improved communication ability. Being able to interact with others, by virtue of being able to convey needs, wants and opinions verbally is obviously an improvement for a previously non-verbal individual. Blackwell et al's point was that the 'whole' measure of the individuals communicative ability was likely to improve. That is, measures of the individuals own residual communicative ability, whether verbal or non-verbal. In addition to increased environmental opportunity Blackwell et al stressed the importance of environmental control, or simply the opportunity to make choices and decisions. Use of an ACD should enable such choices to be made.

SUMMARY

Within this chapter the prevalence of language disorders within a typical LD population has been found to be very high, over 90%. Articulation disorders approaching 25%. Attention has been paid to the communicative effectiveness of typical LD individuals despite their handicaps. That is when considered both verbal and non-verbal behaviour. The impact of age, gender, location, medical condition and degree of physical handicap have also been considered in relation to communicative effectiveness.

Common language and cognitive handicaps have been considered. It has been suggested that language disorders generally occur in parallel with other cognitive handicaps such as poor short term memory, poor sequencing skills, short attention span etc...

Standard classification and epidemiological measures from the DSM-IIIR have been covered for sake of completeness. However, the emphasis has been on functional or adaptive behavioural measurement. For example, measures of how well the individual interacts with the environment and others, or simply measures of what they are capable of, as opposed to measures of what they are not capable of, e.g. IQ test measures.

The relevance of ACD intervention in the light of population characteristics has been discussed. In addition to providing a means of *vocal* communication, potential secondary benefits have been highlighted in relation to individuals cognitive and other handicaps.

The effectiveness of any ACD must take account of *all* handicaps, and not those that are language based in order to be successful for the population in question. The next three chapter will focus on the need for appropriate design of an ACD to match the needs and abilities of potential users. These cover previous devices, potential interface metaphor, and user interface design respectively.

CHAPTER 3: Previous Computer Based Augmentative Communication Devices and Other Systems

Before attempting to discuss one's own attempt to produce a *better* ACD for a LD population than those that preceded it, logic dictates that one should review the generic features of computer based ACD's. That is, to review the forms of speech output, symbol sets employed, user interface methodology, and advanced approaches to ACD implementation.

To illustrate these generic features in practice commercially available ACD's are discussed. Advanced approaches are also considered. Finally future trends are considered briefly.

INTRODUCTION

Augmentative communication systems have undoubtedly been in existence since then dawn of time as a supplant to spoken language. The use of gesture, and in more recent times, sign language are exemplars.

Despite the number of augmentative communication systems that have been and gone over the centuries it is only in relatively recent times that a standardisation of sorts has become apparent. British Sign Language, American Sign Language, Makaton, Bliss and Rebus perhaps typify this standardisation. Unfortunately, many of these systems, despite the promise of offering an alternative communication channel to the spoken word are only available to the more able of the non-vocal, or speech impoverished population. Just as spoken language is an abstract, and indeed an arbitrary system of communication, so too are these alternatives. As highlighted in the previous chapter many of those who might benefit from the use of such systems are handicapped in doing so by cognitive and other impairments. Many find the complexity of learning such systems overburdening and tend to learn only the rudimentary 'symbol set' needed for interaction. Combined with the fact that many such individuals interact with the general speaking population the load they may be under could be at least twice as great.

During the last twenty or so years computer based Augmentative Communications Devices (ACD) have been produced which have attempted to bridge this gap. That is, to bridge the gap between an alternative symbol set and the general speaking population. The complexity and usefulness of these devices has been inextricably linked to the computer revolution, although many of the symbol sets used pre-date many computer systems. Many of the earlier systems, and for that matter the majority of their contemporaries, attempt to turn an alternative symbol set manipulated by the non-vocal user in to synthesised speech which is *understood* by the general speaking population.

Computerised devices of this ilk still search for the 'holy grail' of turning compressed alternative symbol set input into easily understood human like speech. However, despite advancing technologies and a better understanding of user interaction many devices fail to achieve their full potential as in common with the alternative communication systems they might mirror, the device itself adds another layer of complexity. The classic example being of devices which turn typewritten text in to synthesised speech. The obvious failing being that users must have a near normal grasp of written English, be able to type, and to be able to combine both skills at such a speed as to communicate adequately. The cognitive feat required to interact via such devices would defeat many mere mortals, not to mention those intended users, commonly with a mixture of cognitive, linguistic and motor impairments.

In an attempt to address these early failings a range of devices has been produced that in combination have the goal of filling every niche. However, a generalised device of adequate quality still remains a goal for research.

Devices produced to date range from simple 'text-to-speech' devices mentioned earlier, to the more complex featuring prediction, and to the cutting-edge which may encompass voice or gesture recognition. In attempting to review existing devices it is useful to attempt to categorise devices into various sub-types, to highlight the strengths and weaknesses of the generic class the represent, and then to move on to discuss specific commercially available devices.

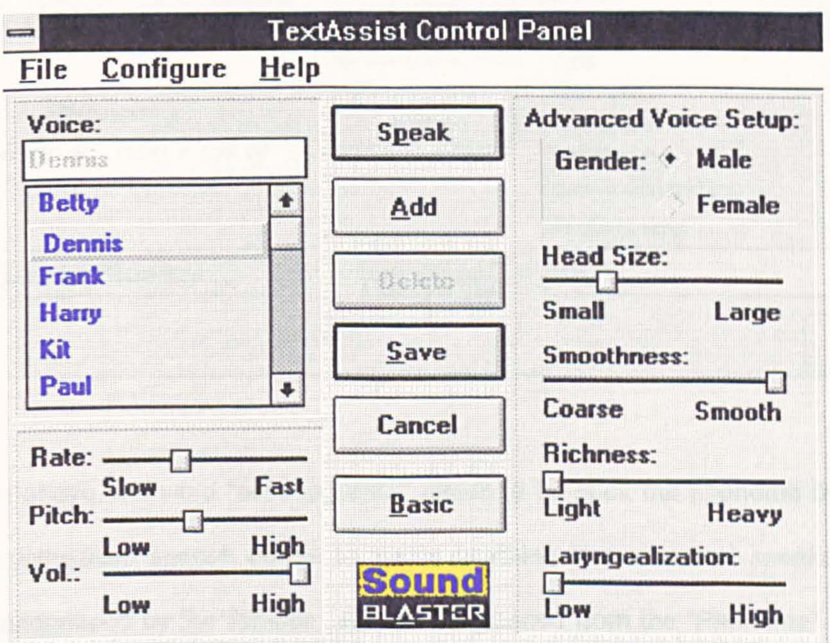
THE FORMS OF SPEECH OUTPUT

Crucial to the success of any ACD is the type and quality of the *speech* output it produces. With current technology output is limited to either synthesised or digitised speech. Both methods have their advantages and disadvantages.

Synthesised speech

The majority of the more potable devices continue to make use of speech synthesis. Improvements in both technology and the algorithms used to generate the actual *speech* have led to the output from such devices becoming more acceptable. That is, in terms of both quality and clarity. Such systems now offer a vast improvement over the monotone robotic sounding systems of the 70's and 80's and offer a range of tuning options, with DECTalk being employed by many systems.

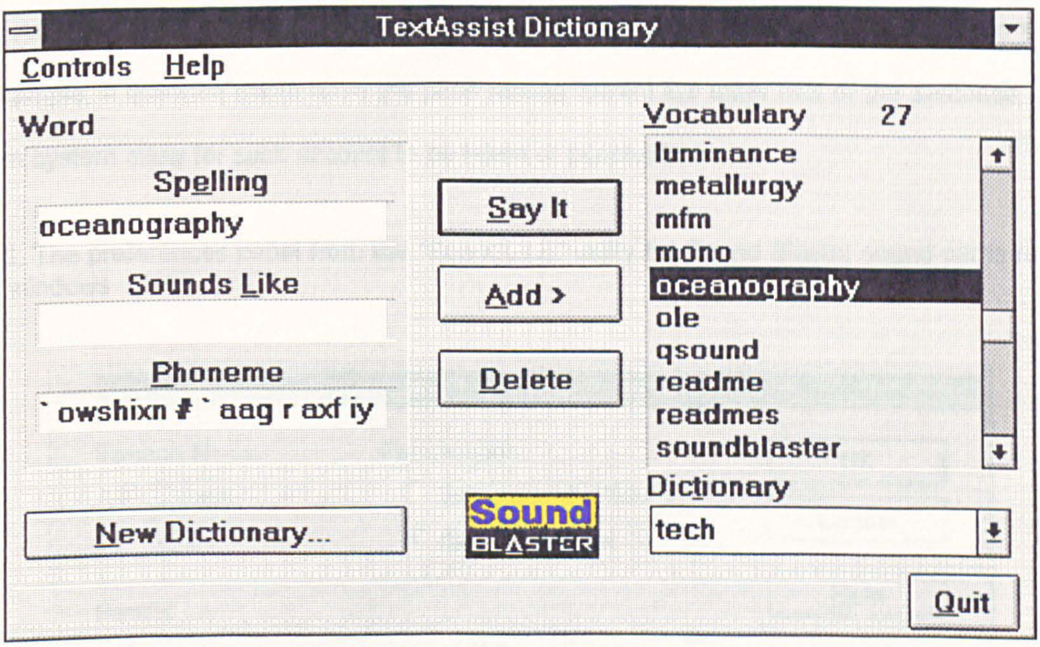
Fig 3.1, The control panel from the "TextAssist" utility for Sound Blaster sound cards running under windows



With modern synthesised speech systems many 'tweaking' options are available. These mean that gender can be catered for, along with options to alter the age associated with the speech. Standard options for rate, pitch and volume can combine to 'tune' the speech to the individual to a remarkable degree.

Such systems still however exhibit similar problems to those of ten or twenty years ago in the respect that they simply translate pure English text into a phonetic script which is then converted into speech. The problem being that not all words translate well when a direct phonetic translation is applied. For example the word "yacht" will generally produce a piece of synthesised speech sounding more like "what". Modern systems however have the capacity to look 'difficult' words up in a phonetic dictionary. For example, in "TextAsisst" technical words can be entered into a phonetic dictionary :-

Fig 3.2, The dictionary from the "TextAssist" utility for Sound Blaster sound cards running under windows



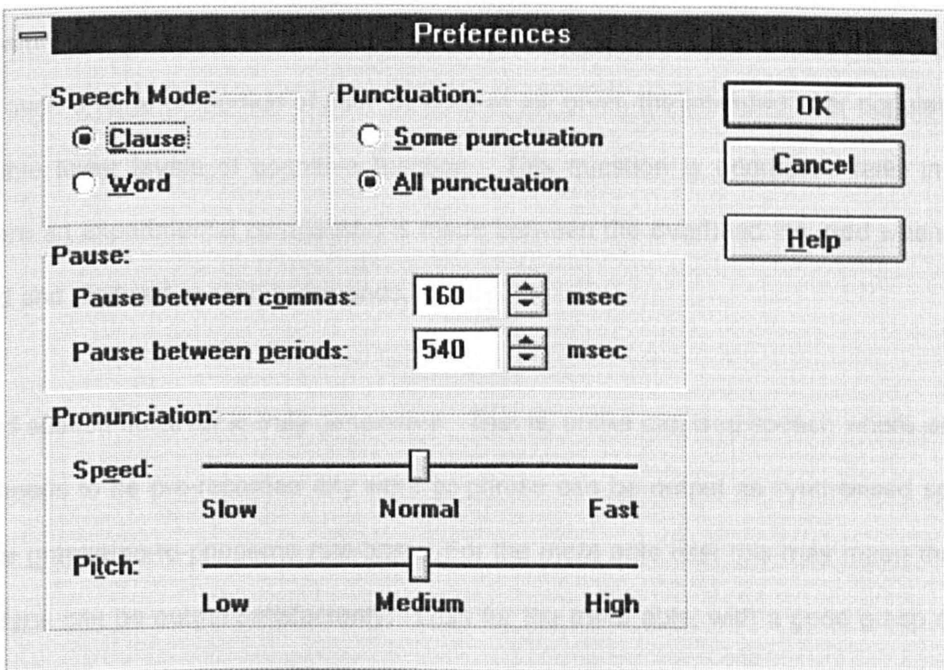
In the example above, the word "oceanography" needs to be spelt out phoneme by phoneme for the resulting synthesised speech output to sound anything like a normal speaker, and consequently to be understood by the listener. As can be gleaned from the "Phoneme" entry box contents :-

` owshixn # ` aag r axf iy

is not a direct phonetic transcription of "oceanography". That is, the translation has to be altered until the speech output more closely approximates the word in question, rather than simply being a straight translation. What may be considered the wrong phoneme set may have to be included to achieve the desired result. Obviously the average user is likely to be unable to achieve such 'tweaks' without help. It is highly unlikely that anyone could actually directly input words in the correct phonetics for the speech flow to have optimum pronunciation qualities. However, this method has been implemented in several devices, the most notable being "FingerFonicks" by Words+ Inc. To what success remains debatable.

Modern synthesis systems also have the ability to take account of intonation, unlike their mono-tone forebears. However, this is unlikely to match the intonnative qualities of natural speech. Some ACD's support extra commands, or symbolics, that allow the tone of an output to be set. For example, a question might have the pitch raising toward the latter half of the sentence. Most modern system allow for such account to be taken of punctuation :-

Fig 3.3, The preferences panel from the "SoundOLE" utility for Sound Blaster sound cards running under windows



As with previous systems words can be spelt and then output letter by letter if required, e.g. :-

Input	Transcription of sound output
y e l l o w	yeh e l l ow w

Modern systems can also deal with conjunctions in a more intelligent way. For example :-

Input	Older system output	Modern system output
£100	pound one zero zero	one hundred pounds

Despite the undoubted progress of synthesised speech it is still a long way from approaching the quality of digitised speech which can be of up to CD recording quality. However, the future does look brighter as there is research afoot to combine the low overhead of synthesised speech with reusable digitised phonetic segments to produce a more natural sounding output. Most current output however is of average quality.

The key advantage of synthesised speech being that it has a low overhead in terms of hardware needed and is simple to implement. This trade-off is based on the quality of the resulting output. In the light of this degradation of output quality as compared with natural speech one could argue that listening to and comprehending synthesised speech incurs a greater cognitive overhead as compared with digitised speech. The leads to the question of whether synthesised speech should be used in such a large proportion of such devices at all, given the intended user population, with their probable lower levels of cognitive function. This question is addressed later in chapter seven, where an experimental comparison is made between the overhead incurred when obeying synthesised and digitised speech commands.

Synthesised speech however is truly generative. That is, unlike digitised speech where each word or phrase needs to be pre-recorded any word or phrase can be output as synthesised speech by virtue of the grapheme-to-phoneme rule-base. For the more able user this may mean that everything they type can be output satisfactorily. Thus for the more able, with a good grasp of written English, speech synthesis may provide a more flexible tool for normal conversation with its inherent richness and diversity. On the other hand for systems that are message based or have a finite symbol set such flexibility is less crucial, and it is suggested in the case of LD user clarity is more important.

Digitised speech

It is only within the last five to ten years that the use of digitised speech has become practicable given the ever increasing power of computer hardware available. However, it is only within the last few years that implementing full scale ACD's making use of has become feasible.

Unlike synthesised speech systems the use of digitised human speech incurs a large hardware and processing overhead. The digitisation process can be alikened to the recording of a human voice using a standard Dictaphone. The longer the speech segment you record the more space on the tape you use up, or in computer terms the more space on the hard disk, or within memory, is filled. The space required to store digitised speech varies according to the sampling quality, or rate and the number of channels used. To illustrate the effect of sampling rate the table below shows the effect of sampling rate on size of resulting digitised speech file for the sampled speech "Hello Richard".

As can be seen from the spectral analysis both the leading and trailing silences have been removed to make the sample as short as possible.

Fig 3.4, The effect of sampling rates on size of resulting digitised speech file for the digitised speech "Hello Richard"

Original 8 bit 11 KHz sample which is 0.73 seconds in length

Mono / Stereo	Bit resolution	Sample rate in KHz	Length in Kbytes per sec required for storage
Mono	8	5.5	5.4
Mono	8	11	7.8
Mono	8	22	15.6
Mono	8	44	31.3
Mono	16	5.5	7.8
Mono	16	11	15.6
Mono	16	22	31.3
Mono	16	44	62.6
Stereo	16	44	125.2

The quality of output then is dependent on both the sampling rate and the bit resolution. The higher the quality of the output the larger the storage space needed on the computers storage system, typically hard disk or EPROM.

As a rule of thumb a sampling rate of around 11 KHz is suitable for voice recording, 22 KHz is suitable for tape quality recording, and 44 KHz is suitable for CD quality recording. A bit resolution of 8-bit data gives a lower sound quality comparable to that of a cassette tape, whereas 16-bit data gives the highest sound quality comparable to that of a CD.

Storage requirements usually follow the formulas :-

Mono 8-bit storage requirement = 1 x Sample rate x Number of seconds of recording

Mono 16-bit storage requirement = 2 x Sample rate x Number of seconds of recording

If stereo is required the storage requirement is doubled again. However, thankfully for speech a resolution of 8-bits with a rate of between 5 KHz and 11 KHz is sufficient to offer very good quality output. This translates in to a storage requirement of about 10 Kbytes per second of speech. Applying this rule and assuming that you could record one word per second a storage table might look something like :-

Fig 3.4a, A hypothetical storage table based on sampling rate

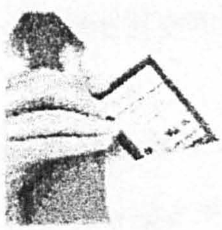
No of unique words	Recording time	Storage space required
60	60 secs (1 min)	600 kb (≈ 0.6 Mbytes)
600	600 secs (10 mins)	6000 Kb (≈ 6.0 Mbytes)
3600	3600 secs (1 hr)	36000 Kb (≈ 36 Mbytes)

Digitised speech of the quality mentioned offers tape recorder quality and can carry all the intonation of the speech it is a facsimile of. In addition use of digitisation offers the facility to record not just speech but also other sounds and music. Something which synthesised speech will *never* be able to offer. Digitised human speech is also inherently natural and is more likely to be understood easily.

Despite digitised speech seeming to offer many advantages over synthesised speech its major drawback is one of its lack of generalisability to words which have not been previously recorded. New words or phrases would need to be recorded as and when required. This is one of the great strengths of synthesised speech. In addition for sake of consistency one would prefer for all speech to be sampled from the same person. In synthesis systems this is not required.

One of the few remaining objections to digitised speech is the storage requirements, and the processing demands it paces on the hardware. However, with more powerful processors and high capacity miniature hard disks a truly portable ACD is no longer a fantasy :-

Fig 3.5, A Calluna pocket sized hard disk capable of storing up to five hours or more of digitised speech



Offering 130 Mb of storage and measuring just 85 x 54 x 10 mm with a weight of around 85g.

Theoretically such a drive could store over 3.5 hrs of digitised speech or around 12,600 individual digitised words.

Such developments are likely to continue with the increasing demands of software. For example, the word processor I am currently using takes up around 30 Mb, or, enough for an hours digitised speech. Perhaps indicative of future trends are the 32 Mb PCMCIA II cards offering 32 Mbytes of storage space within a package no larger or thicker than a credit card.

With modern equipment and software it is also possible to incorporate the pseudo-generative characteristics of synthesised speech and therefore reduce storage requirements drastically. That is, not to store whole set phrases but to store individual digitised words which could then be linked to form a phrase or sentence. Indeed it is this method with the author implements in Easy Speaker for speech storage. An example of such usage can be heard when telephoning directory enquiries, where the number required is read back by linking digitised speech recordings of a phrase and then individual digits. With the advantages afforded by use digitised speech, combined with the reducing demands on equipment one might expect ACD's to reflect this shift in technology and availability.

One of the main disadvantages of this approach, when applied to digitised speech, is that when discrete sections are combined it can often sound unnatural, as each section may have been taken from an inappropriate context. This problem of concatenating discrete speech can be overcome to some extent by partially preconstructing a series of useful messages. Users can then slot in the required word(s) with little change in the overall naturalness of the output. Unfortunately this is not a complete solution because as the individuals communication improves large amount of predefined phrases are needed.

Given that devices are now capable of producing high quality *spoken* output, the fundamental question remains of how best to access a potentially massive vocabulary ?

COMMON SYMBOL SETS AND THEIR IMPLEMENTATION

Having discussed the forms of output available to ACD designers the overriding issue is then one of how to represent the stored speech in a way that the user can understand and manipulate easily. This is usually done by means of an alternative symbol set whereby each symbol represents either a word or phrase. The finite details of symbol sets are unimportant, what is are the rudimentaries of their implementation, and possible consequences.

Text-to-speech

The first and most obvious symbol set to augment spoken language is the written word. This has been taken literally and has been transformed into "text-to-speech" ACD's. Where as the name suggests plain English text is turned into synthesised speech output. This is usually done in a manner similar to that described above when discussing speech synthesis. However, there is no reason why digitised speech could not be used given a sufficient digitised vocabulary. To date however most systems have translated plain English text into synthesised speech. Differences between devices are small with usually the only difference being the interface used to *enter* the text.

On many occasions the interface for creating text has been the QWERTY keyboard, or a close derivative. Obviously this makes assumptions about the user and may place a great demand on them. It assumes that the user has a good grasp of English, has the ranges of movement required to type and can do both at such a speed as to communicate effectively. Something all but the most able cognitively and physically can achieve. Alternatives to interactive discourse include methods where pre-generated text files are read back via a text-to-speech algorithm. One of the more complex interfaces is that implemented by Stephen Hawking's synthesiser where by one button click is all that is needed to generate the text needed for subsequent output. This is, done via a compressed input interface. A scanning horizontal bar passes over a grid of words. When done via a compressed input interface. A scanning horizontal bar passes over a grid of words. When the row containing the target word is reached, a button press starts a vertical bar scanning which then passes over that row. When the button is pressed that word is added to the construction currently underway :-

Fig 3.6a, A bar scans horizontally until a button is pressed

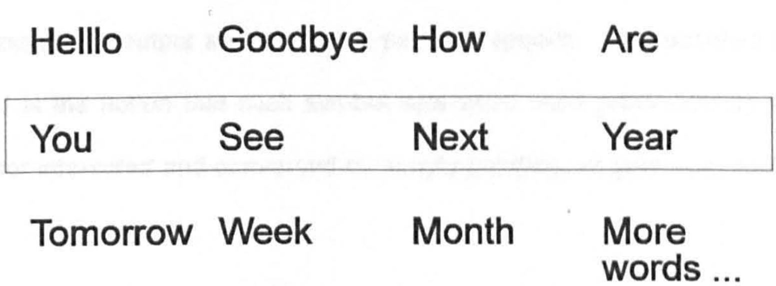
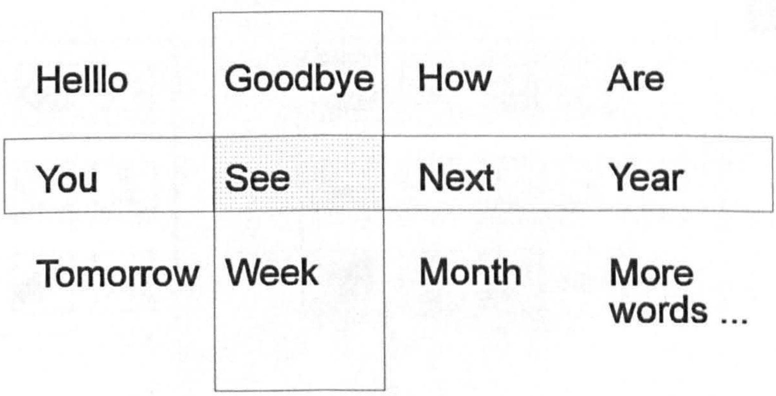


Fig 3.6b, A bar then scans vertically until a button is pressed



In the example above the word "See" has been selected by two single button presses.

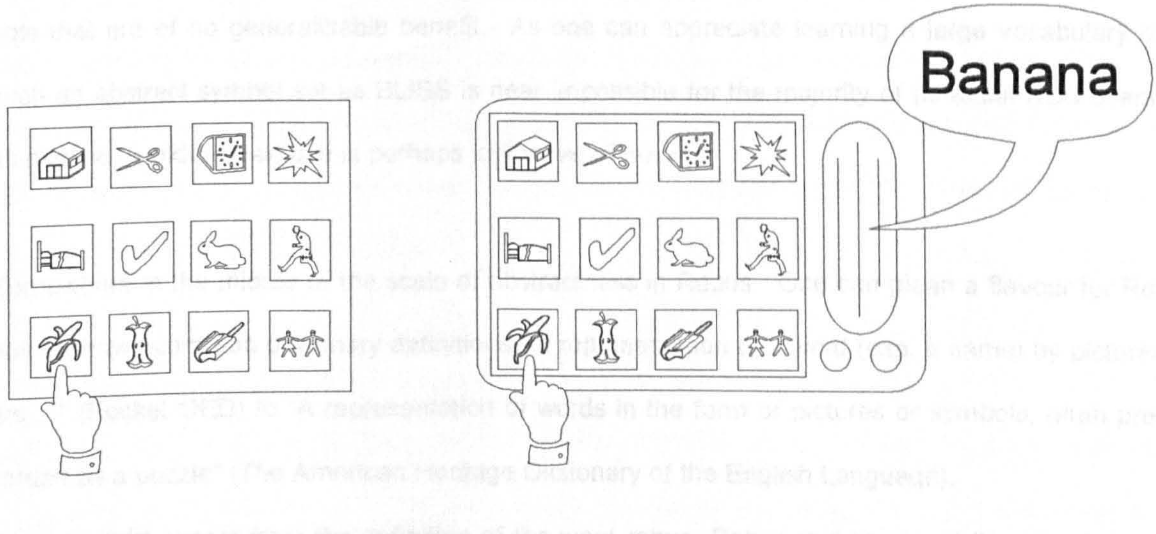
As the majority of ACD users, and potential ACD users, do not possess the required levels of ability many alternative symbol sets have been constructed and bolted on to ACD's. These symbol sets range from the more abstract such as Bliss Symbolics, to the more pictorial Rebus. Their basic task being to provide access to an underlying text-to-speech, or digitised speech vocabulary.

Symbol based systems

Symbol based systems attempt to provide an alternative interface to ACD's other than the textual type of interaction described earlier. Many of the symbol systems used predate the invention of ACD's and have been used as alternative communication strategies in their own right.

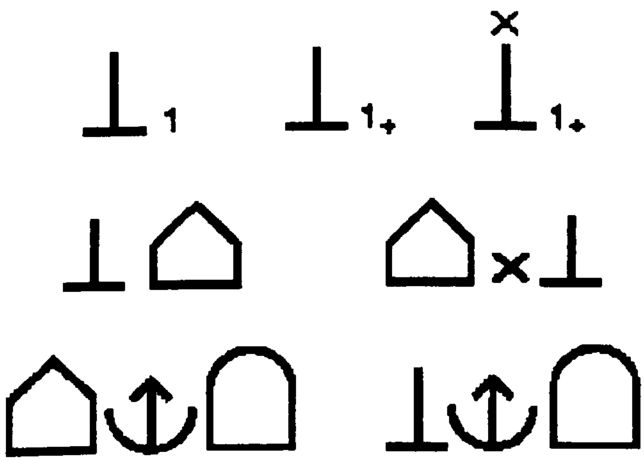
As the majority of these systems are iconicity or pictorially based they are generally integrated into the interface of an ACD by means of superimposing the particular graphic onto a given 'button' which when pressed will output an associated piece of speech. The principle on which this approach is based is the notion that such symbol sets were used previously on form type boards. Whereby the user interacted and conversed by simply pointing, or gesturing, to a given symbol, or symbols.

Fig 3.7, A simple paper and card based formboard approach (left) translated into a common ACD (right)



Examples of typical symbol based systems are Bliss and Rebus. Such symbol systems range from the abstract to the more concrete. With the more abstract systems needing to be learnt almost by route, and generally available only to the more able.

Bliss Symbolics for example requires that an arbitrary symbol set be learnt by the user. The symbol set itself follows a quite complex logic that must be learnt in order to form new conjunctions. In the example below one can quite easily see how a related symbol set is constructed by adding and subtracting other features to the symbol in question :-



Top row (left to right): I, my, our
Middle: resident, residential institution
Bottom: school, teacher

BLISS despite having been used with some success with certain groups has proved to place too much of a burden on the majority of users to be of widespread benefit. This is especially true of the Learning Disabled population who may spend a disproportionate period of time learning symbols that are of no generalisable benefit. As one can appreciate learning a large vocabulary of such an abstract symbol set as BLISS is near impossible for the majority of potential ACD users. Its decline in widespread use is perhaps indicative of this.

Somewhere in the middle of the scale of abstractness in Rebus. One can glean a flavour for Rebus from two common dictionary definitions, "a representation of a word (esp. a name) by pictures etc... " (Pocket OED) to "A representation of words in the form of pictures or symbols, often presented as a puzzle" (The American Heritage Dictionary of the English Language).

As one might expect from the definition of the word rebus, Rebus makes use of line drawings to represent words and actions. Using a predefined symbol set Rebus symbols attempt to 'act out' the word or action the symbol represents. For example, a line drawing of a plate of food might signify that a the user would like some dinner. An example of Rebus symbols are shown below :-

Fig 3.9a, An example of rebus symbols

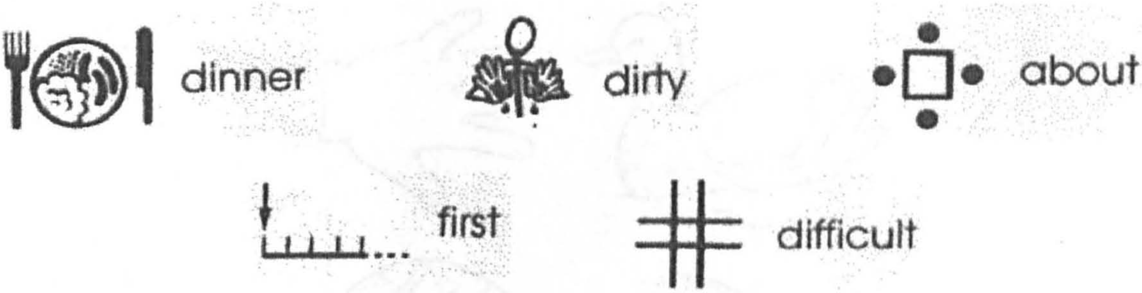
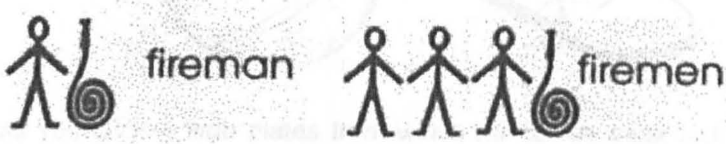


Fig 3.9b, The additive nature of rebus symbols

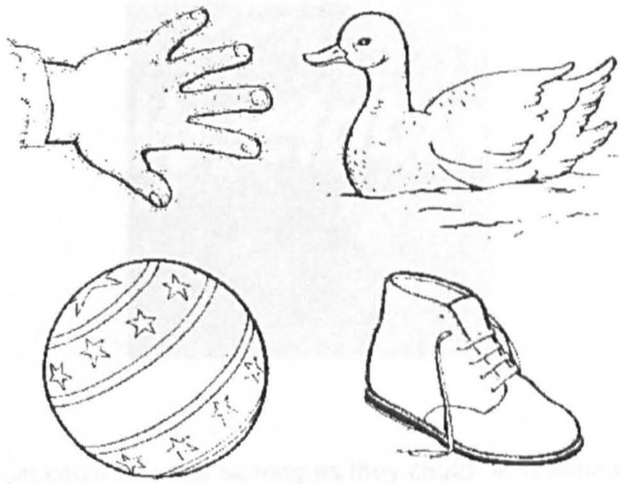


Using a less abstract symbol set such as rebus is initially quite appealing. Hurlbut, Iwata and Green (1982) compared Bliss and more iconic Rebus-like symbols. They found that the more iconic symbols excelled in measures of acquisition, generalisation, maintenance, and spontaneous use when used with a learning disabled group. Further support is lent by Clark (1981) who compared recall rates for Bliss, Rebus and printed words among pre-schoolers. Again the more concrete symbols proved to be superior.

In support of using such a symbolology several clinical tests for children and learning disabled individuals use such an approach due to the limited responses needed. Indeed the author used one such test, the British Picture Vocabulary Scale (BPVS) for assessing users general levels language of ability before starting Easy Speaker intervention.

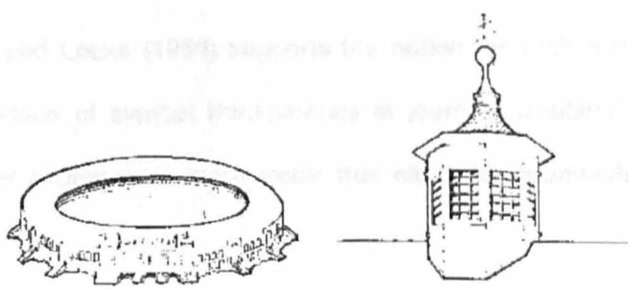
Such tests make use of a highly standardised set of iconic symbols which have been standardised as having a specific meaning. Symbols range from the very simple to the highly complex. Testees are asked if they can indicate a good example of a word the administrator has just spoken. In the example below item one from the BPVS illustrates the point :-

Fig 3.10a, Item 1 from the BPVS, where the testee is asked to point to a hand



The BPVS has 150 such 4 item plates from which the testee must pick the best example of a given category. To give an example of how abstract later items on the test can be requested items 75 and 150 are shown below :-

Fig 3.10b, Requested items 75 (stadium) and 150 (cupola) from the BPVS



The alternative to using a recognised pictorial symbology is to use one that is either generalisable or idiosyncratic to the user.

The approach which Easy Speaker uses is somewhere in the middle-ground as it uses real digitised photos. That is, images that are of passport size colour photo quality. For example, the digitised words "breakfast cereal" might be represented pictorially be a picture of a real box of "Weetabix" :-

Fig 3.11, A box of "Weetabix" signifying the digitised speech "Breakfast Cereal"

If required any symbol set could be used so long as they could be scanned in with the appropriate hardware.

The goal is to move from abstract systems to those of a more concrete nature, such as that used in Easy Speaker. Concreteness is taken to its ultimate conclusion in the "Adaptech Voicepal" which uses real objects which when touched produce digitised speech.

Research by Mirenda and Locke (1989) supports the notion for such a move to more concrete systems. In a comparison of symbol transparency in learning disabled subjects Mirenda et al found that found colour photos were more iconic that either black-and-white photographs or line drawings.

Reichle, Sigafoos and Remmington (1991) note how it is "... more practical to take advantage of any symbol systems which already operate in a learner's environment". They sight the example of learners who have already been exposed to many incidental parings of a product logo with the receipt and consumption of the product itself. Such pairings, through conditioning could rapidly learn to use such a symbolically based system.

The most concrete symbol system, is that which is idiosyncratic to the user, but also generalisable. Namely pictures in which they are seen interacting with the environment. Perhaps in-line with the example above a picture on the user holding a box of breakfast cereal would be highly concrete to them. Such a picture might also be generalisable to other users due to the main focus in the picture being a box of Weetabix and not the person serving them. The hope being that cognitively impaired learning disabled users would benefit markedly from such a concrete approach.

For the more able the expression of abstract ideas and concepts through a concrete symbol set is problematic. For example, it is difficult to have a set of universally understood symbols for abstract ideas such as love and hate. With ingenuity this might be accomplished by using idiosyncratic symbols unique to a given user. Fortunately, because Easy Speaker uses digitised images it is possible for any symbol set to be used either in isolation or in combination. Thus a combination of strongly representational symbols (digitised pictures) could be used alongside a small BLISS vocabulary which being an abstract system could be used for the expression of abstract concepts not readily expressed by a completely concrete system. It is also possible that many LD users would not use Easy Speaker for the expression of abstract concepts anyway preferring to use it as a simple but effective concrete messaging system to aid in communicative effectiveness. Some LD individuals express abstract concepts concretely, e.g. they will hug someone they are pleased to see rather than tell them.

Given the variety of symbol sets that are available the next issue is how to transform them into a usable user interface ?

USER INTERFACE METHODOLOGY

Many of the forerunners of computer based ACD's were simply based around a user pointing to a symbol or symbols in turn shown on a form board in order to communicate their needs and wishes. This technique has been transferred directly to many computer based ACD's, as highlighted above. Many commercially available devices have a set of symbol based cards which act as overlays a standard touch sensitive formboard. When these symbols are touched output is generated, whether it be synthesised or digitised speech. The restriction with such an approach is obvious: only a limited symbol set can be displayed at any one time before another overlay need be used and the device reset to recognise the new symbol set.

Text based systems obviously do not suffer from such a drawback, as they can either be based on an alphanumeric keyboard or use a scanning methodology such as that discussed earlier. Overlay based devices have attempted to circumvent this problem by one of several methods.

Firstly they have either made the symbols smaller so that a larger number can be packed into the same area. Again this presents problems as users tend to have poor motor skills and many have uncorrected visual impairments which have gone unnoticed. For example, the commercially available Liberator by Prentke Romich has no fewer than 128 pictorial symbols in a package no larger than a sheet on A4. Each symbol being approximately 1 cm square, or around the size of a postage stamp. Obviously such a high symbol density places many extra demands on the user.

Furthermore, if the user needs to make use of one of the many alternative forms of input aids, e.g. a head pointer, the task of symbol selection becomes another layer more complex.

Another alternative based upon a similar principle has been to make each symbol represent more that one utterance. This is analogous to a standard keyboard where shift is held down to access a second set of symbols shown over the numeric keys, e.g. the "5" key when hit with shift down produces a "%" and so on. Again this adds to the complexity of the system as a whole, and may defeat many potential users, especially if the symbols are already densely packed. Even with prolonged practice one might expect cognitively impaired users to never achieve their full potential.

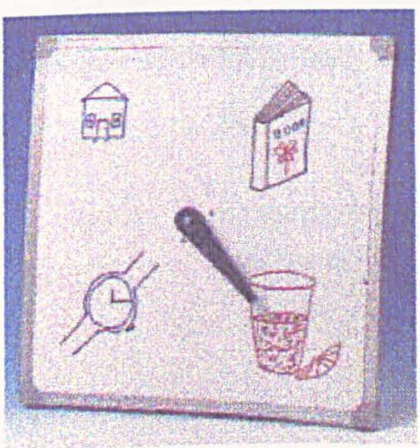
Newer systems tend to be based more on standard computer equipment and can make use touch screen technology. Although many such systems are still confined to the lab they offer the possibility of achieving the same goal as overlay based approaches but with the flexibility of possible access to multiple 'overlays'. Many have termed such systems as being dynamically display based. Typically systems have represented multiple overlays by means of a forward and backward button which enables overlays, and thus symbol set, to be 'virtually' changed on screen. Thus the working vocabulary is enlarged. An example of this arrangement would be Chute's, Apple MAC based "Speak Easy" which uses Rebus like symbology combined with synthesised speech (Chute, Conn, Dipasquale and Hoag 1988).

Logic would suggest that what is needed is a radical rethink of interface so as to enable users access to a large vocabulary at little extra cognitive effort. This has hopefully been achieved with Easy Speaker, which allow access to a large symbol based vocabulary by means of a hypermedia metaphor. That is, certain symbols instead of representing digitised speech represent links to other 'overlays' of symbols. Thus the user can navigate through a large vocabulary with relative ease. Such an interface methodology is hoped to be of benefit in terms of cognitive training and enhancement also. The rationale for such an interface and it's possible benefits are discussed more fully in chapter four, as is the tangible evidence that user can make use of such a system in chapter nine.

A Selection of Commercially Available Hardware Based ACD's

Commercially available ACD's range from simple devices costing hundreds of pounds to the more complex costing several thousands. The more basic offer limited communicative possibilities and tend to be aimed toward the less able, whereas the more complex are usable by only the most able. In reviewing a representative selection of available ACD's it is useful to move from the basic to the more complex. In addition to complexity two distinct lines of ACD pedigree are now emerging, namely hardware based and software based. That is, dependent on bespoke hardware to function, and independent of hardware respectively. In this respect it is useful to review generic classes of hardware based ACD first and then move on to the newer software based alternatives. Typically the latter tend to be more advanced anyway.

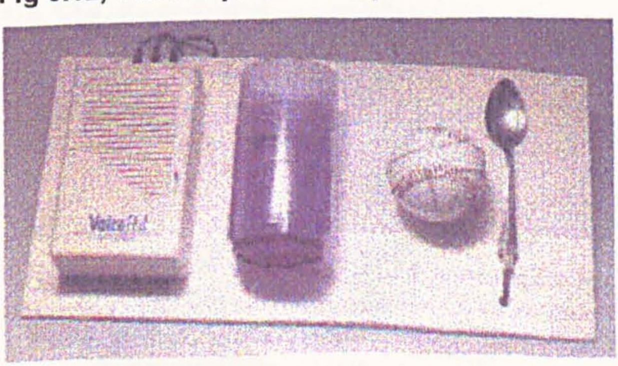
Fig 3.12, The Rotary Indicator



Although the Rotary Indicator doesn't produce any form of *speech* output it is illustrative of how the apparently simple can be used to greatly enhance an otherwise isolated individuals communicative ability.

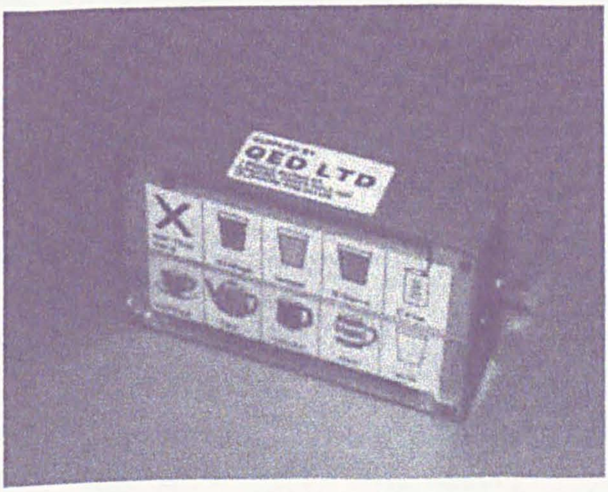
The rotation of the pointer can be controlled via numerous methods, e.g. click start, or click stop etc... Thus by controlling the pointer the user is able to indicate needs and want that previously may have gone unnoticed. The other illustrative aspect of such a simple device is that it allows for any symbol set to be used, e.g. drawings, photos, symbolic vocabularies etc...

Fig 3.12, The Adaptech Voicepal



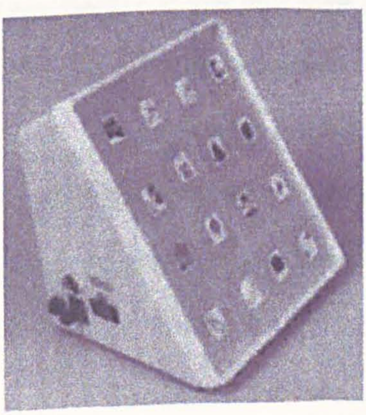
The "Adaptech Voicepal" is based upon a concrete a symbol set as is possible, the actual objects themselves. The device allows for up to five messages to be digitised and replayed when the object in question is touched. Up to 20 seconds of total recording time is available. For example, if the user touches the cup a piece of digitised speech might convey the message "I'm thirsty, can I have a drink ?". Obviously both this and the previous system are aimed at making a real communicative difference for the less able.

Fig 3.13, The Portable Aid for Communication



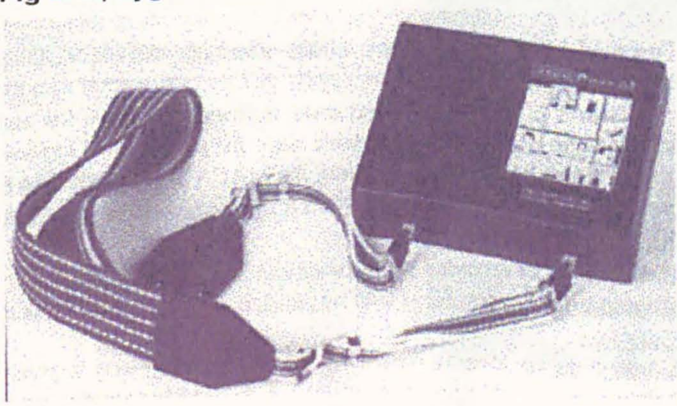
Although the PAC cannot produce speech output it is illustrative of a line of ACD's designed for the slightly more able. The device can accept various cards which are representative of the user needs. These can either be of set symbology or idiosyncratic. To indicate a need the user clicks a single microswitch so that a light moves in a clockwise direction around each symbol. On which ever symbol the lights stops, the *agreed* meaning is then understood by the *listener*.

Fig 3.14, The Teaching/Communication Aid



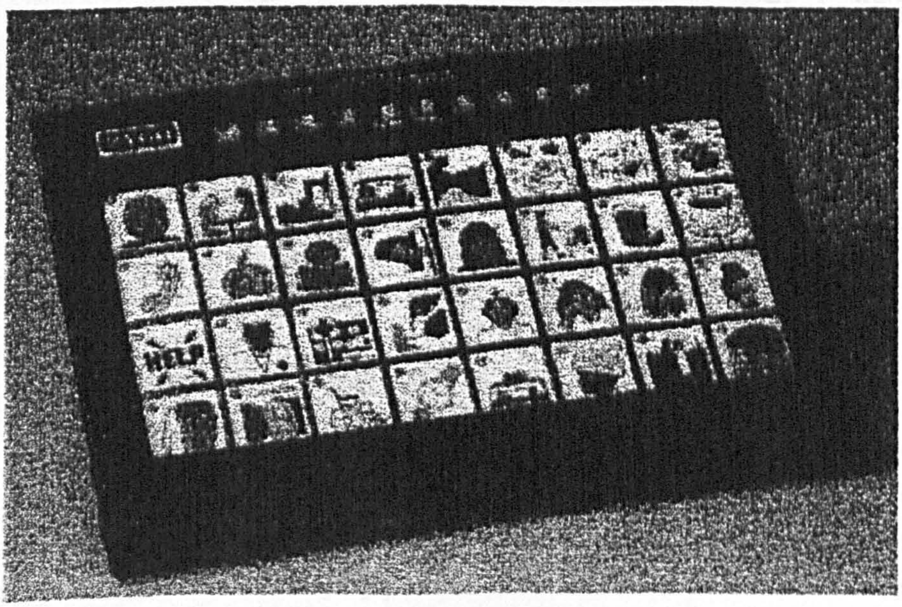
The T/CA is a slightly more advanced version of the PAC. A light can either be stepped around the 16 symbols or can be advanced in a similar way to the PAC. The advantage of the T/CA is that is can be connected to another device, such as the "Zygo Parrot", so as to offer digitised speech output.

Fig 3.15, Zygo Parrot



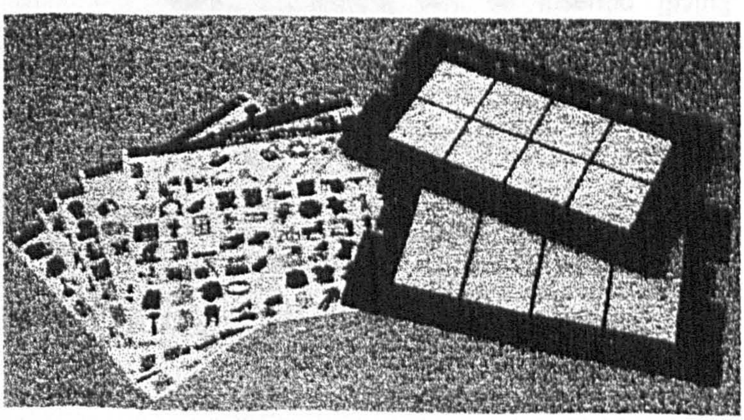
The "Parrot" operating on the standard form board principle can store up to 16 digitised words or phrases which can be output when the appropriate symbol is pressed. 32 seconds of high quality speech can be recorded or this can be extended by using lower quality sampling to 64 seconds. Again any agreed symbology can be used. In common with most devices a scanning type interface can be used.

Fig 3.16a, The Zygo Macaw



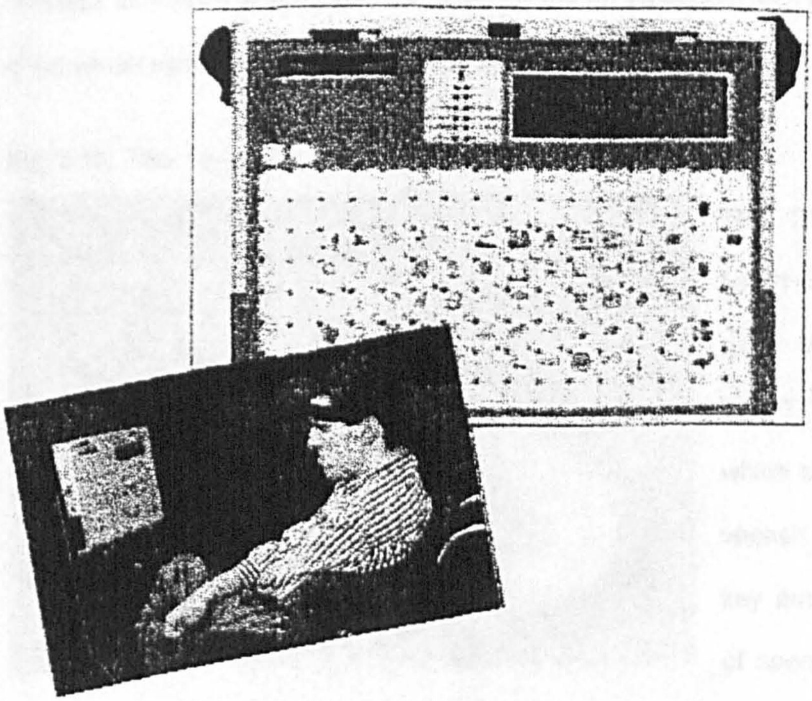
The Zygo Macaw is one of the most advanced formboard based ACD's on the market offering in it's top specification 4.25 minutes of high quality recording, or 8.5 minutes of lower quality speech. It can either be controlled via a variety of scanning methods or by the normal touch symbol mode. The device offers advanced features such as "string mode", "levels mode" and "key link mode". In string mode a longer utterance can be constructed by pressing symbols in turn before a talk symbol is pressed and the speech is output. In levels mode up to eight sections of digitised speech can be linked with each symbol. Using the equivalent of the "shift" key on a normal keyboard any one of these eight can be selected for that symbol. The final mode, key link, allows the speech in up to three symbols to be linked to one symbol. Thus when the linked symbol is pressed the three linked pieces of speech will be output in order.

Fig 3.16b, The Zygo Macaw's interchangeable inlay cards



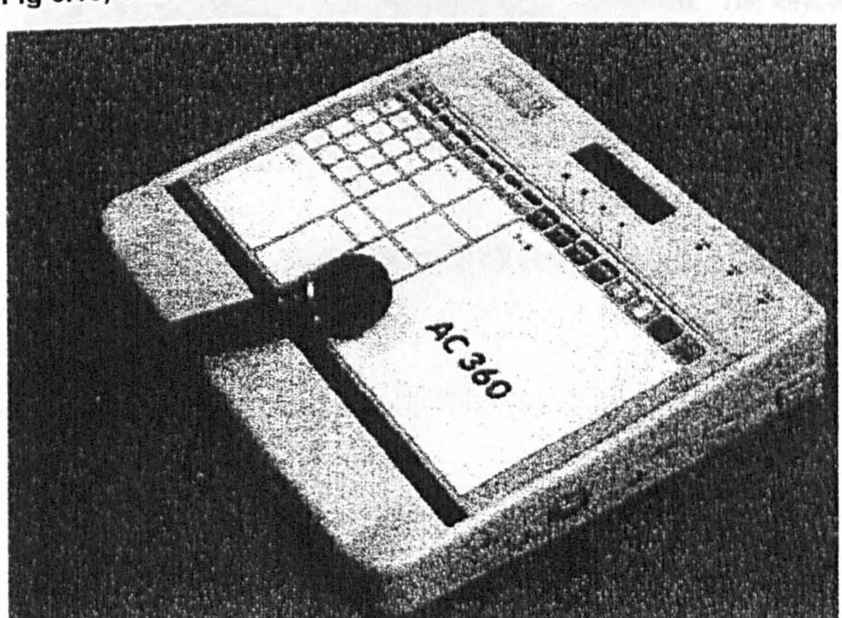
Although symbols are limited to 32 per board other symbol cards can be inserted either by the user or by a carer.

Fig 3.17, The Liberator



The liberator combines a standard formboard approach with a speech synthesiser which can translate the many graphics into synthesised speech. In addition the device can be controlled via either a scanning methodology or via an optical pointer. The underlying text associated with each graphic can also be modified to potentially allow for complete customisation. In common with other devices a 'shift' style mode can be operated whereby each symbol can be associated with more than one underlying piece of text. This method known as "semantic compaction" enables one symbol to represent more than one utterance at the same time based upon the symbols physical and semantic connotations. For example, a symbol of an orange may in combination with other keys represent, "orange", "hot", "sun", "fruit" etc... Semantic compaction has a similar effect to increasing the symbol density. However, some have argued that such multiplicity of meaning can hinder the user by placing them under too much load when operating in this mode. Generally semantic compaction is of most use to the more able.

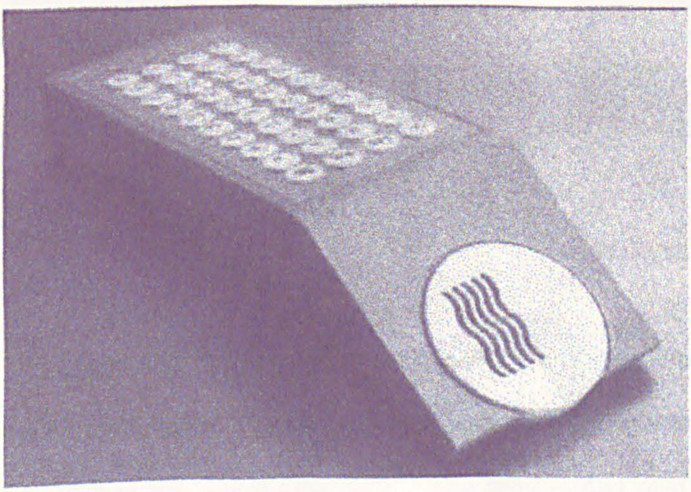
Fig 3.18, The AC360-FX/SX



The "AC360" represents the level of advancement formboard approaches have reached in recent years. In common with the "Macaw" symbol cards can be inserted giving access to a large digitised speech vocabulary. The volume of speech can range from 36 to 72 minutes. In common with the "Macaw" both "String" and "Key link" modes are supported.

Perhaps indicative of the future for such formboard based devices is the inclusion of a 3 1/2" disk drive which enables transfer of material to other machines.

Fig 3.19, The Tiny Talker

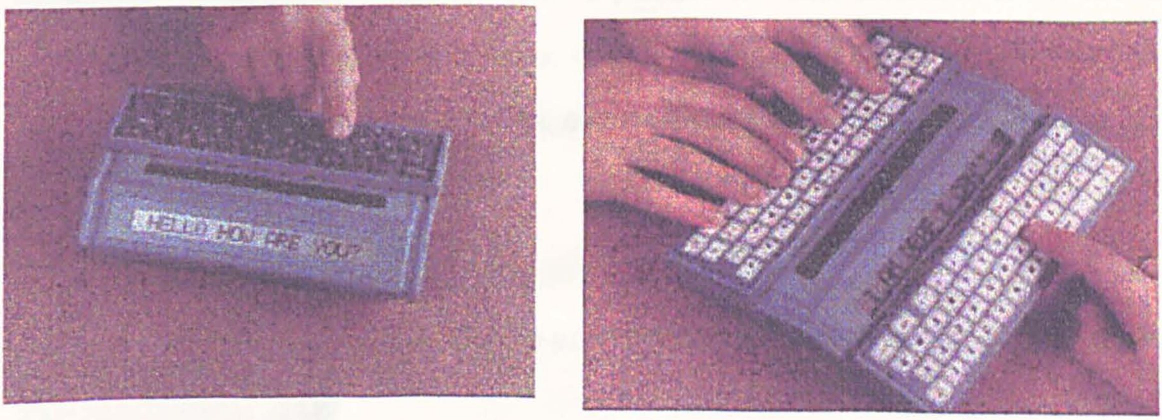


The TT is basically a larger form board dramatically reduced in size so as to achieve hand-held proportions. The TT has an array of 36 buttons which can give access to 72 digitised speech messages by use of a 'shift' key arrangement. Up to 72 seconds of speech can be recorded. Buttons can either be pressed directly or the device can use a scanning method. Speed of use is paramount with one hand operation being usual.

Fig 3.20, The Zygo Secretary

The Zygo Secretary is a hybrid being basically an alphanumeric keyboard controlled device which outputs to a small in-built printer and a Zygo Parrot combined. The keyboard can be used to type message which can then be printed and given to the *listener*. These printed messages can be enhanced or supplanted by activation of digitised speech phrases of up to a combined length of 64 seconds using low quality recording.

Fig 3.21, The Lightwriter



The Lightwriter is analogous to a moving LED 'ticker-tape' type display seen in many shop windows, where what is typed by the user scrolls from right to left, and is seen as an advancement to the likes of the "Zygo Secretary". The listener can also see what is typed as the device has a rearward facing mirroring what the user sees on his screen. Up to 36 user defined phrases can be stored and recalled later. An optional plug in speech synthesiser can be used to reproduce any typed text as *speech*. A scanning version with larger user screen is also available (Lightwriter SL8). The listener screen remains the same size however.

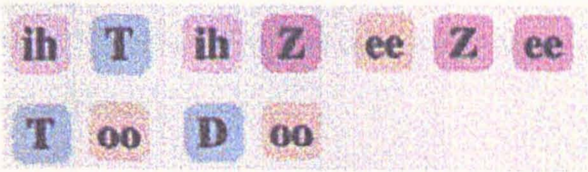
Fig 3.22, The Zygo Talking Notebook

The Zygo Talking Notebook is basically a small laptop computer which can turn text into speech by means of a speech synthesiser. To speed up interaction a system called "ABEX" has been implemented which allows abbreviations to be used to represent user defined phrases. For example, "G1" might reproduce the synthesised speech "Good morning", whereas "G2" might produce "Good afternoon" and so on.

Fig 3.23a, FingerFoniks

"Finger Fonicks" is probably the least accessible of the devices available today, despite offering the potential of being the most adaptable. The goal for "FingerFoniks" is for able users to be able to generate synthesised speech phonetically at conversational speeds.

Fig 3.23b, The phonetically constructed message "I this easy to do"



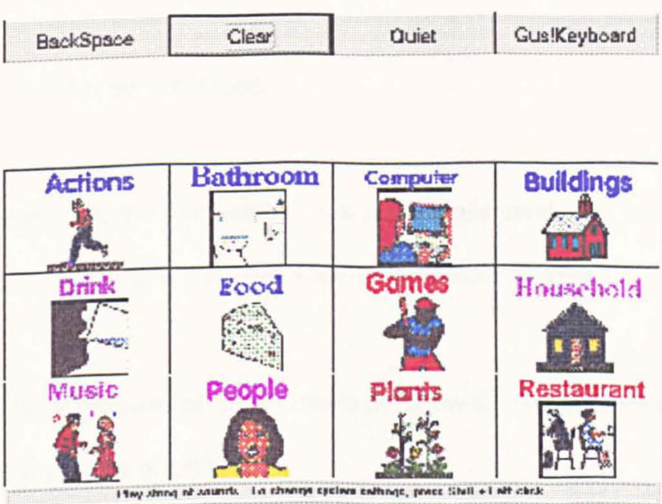
However, if this is unachievable the device can store pre-keyed synthesised speech. "FingerFoniks" is a phonetic synthesiser as opposed to the more common text-to-speech variety. Obviously by virtue of this the device is only accessible to the highly able.

A Selection of Commercially Available Software Based ACD's

Software alternatives to hardware based ACD's have only become viable in recent years. Today many standard machines come 'multimedia' equipped and are easily capable of reproducing digitised or synthesised speech as well as implementing such 'Sci Fi' features as voice recognition.

Despite the availability of numerous hardware platforms the main thrust has been toward IBM PC compatible, and to a lesser extent Apple MACintosh based systems. Many of the available ACD software packages available offer very similar features to those of their hardware counterparts, albeit in some cases 'virtually'. Virtual in the sense that symbol cards are no longer slid in to place but that screens are updated to reflect a change.

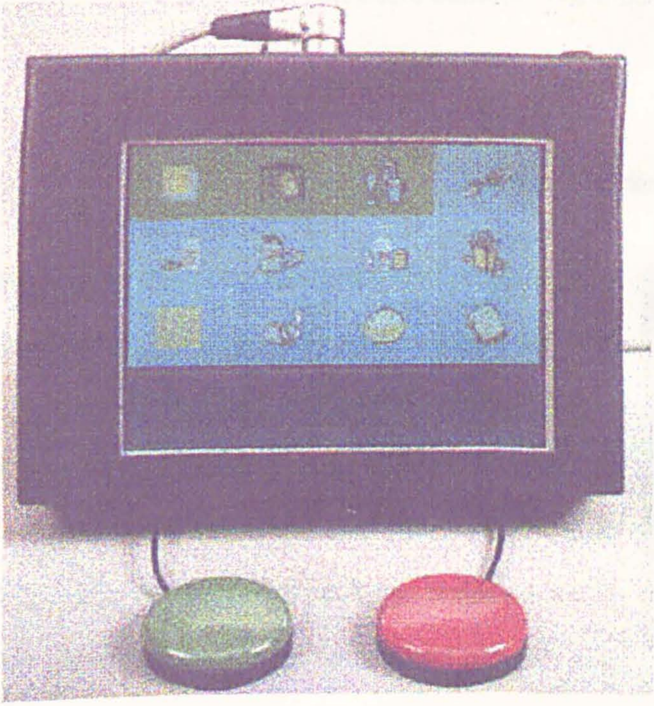
Fig 3.24, The Gus! Multimedia Speech System



The "Gus! multimedia Speech System" makes use of either, or both, synthesised and digitised speech. This speech is represented as symbols which can then be assembled to form longer utterances. Symbols can also represent longer utterances or digitised music or other sounds.

Each screen, or page, can contain anywhere from 2 to 72 such symbols. Users can navigate between pages by using special buttons which act as links to other such pages. The system is capable of producing CD quality digitised speech and offers synthesised speech in male or female English, French, German, Italian and Spanish. The Gus! software will run on any IBM PC compatible running Windows and with an appropriate sound card.

Fig 3.25, The Cameleon



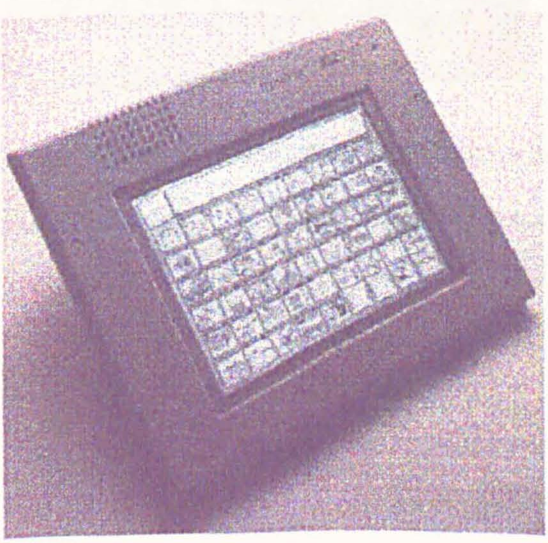
The "Cameleon" operates on much the same principle as the "Gus!" but is aimed at less sophisticated users. The "Cameleon" makes use of "Words+" augmentative communication software and generates synthetic speech from a Multivoice (portable DecTalk) synthesiser.

The "Cameleon" has a built in touch screen which makes interaction more natural. In addition a scanning technique can also be employed.

Although the "Cameleon" is a specifically produced hardware and software combination, the hardware itself is a standard 486slc with colour display.

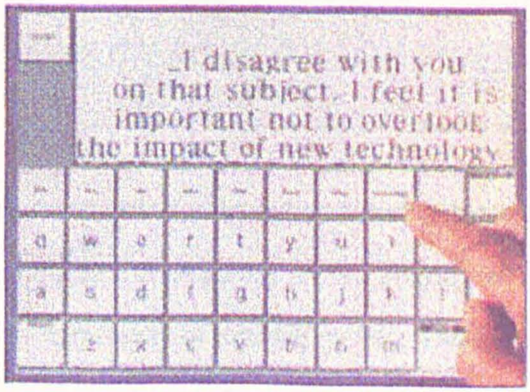
Other features of the "Cameleon" allow it to act as an environmental control, e.g. to switch lights or TV sets on or off etc...

Fig 26a, The DynaVox



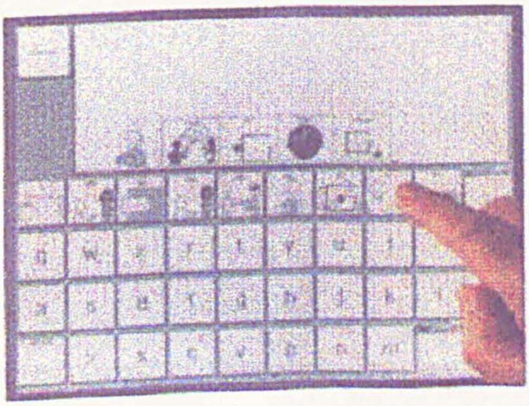
The "DynaVox" in common with the "Cameleon" makes use of synthesised speech via a DecTalk synthesiser driven by either symbol input, textual input, or a combination of the two. The "DynaVox" can also make use of a small amount of digitised speech and music. In common with most other software based devices a touch screen is utilised. If required a scanning methodology can be used, as can other alternative forms of input, e.g. Joysticks, suck'n'blow etc... Anywhere between 1 and 60 hand drawn monochrome symbols can appear on each screen.

Fig 26b, The dynavox in text-to-speech prediction mode



For more able users the "DynaVox" offers a pure text-to-speech mode within which users input text via a virtual keyboard provided on screen. One of the advanced features of the "DynaVox" is that it offers word prediction features. As words are constructed a 'hot-palette' of possible completed words are shown. The user is free to select any of these selections or can continue typing hoping for more accurate predictions or until they manually complete entry of the word.

Fig 26d, DynaVox can utilise a combination of text and symbol input



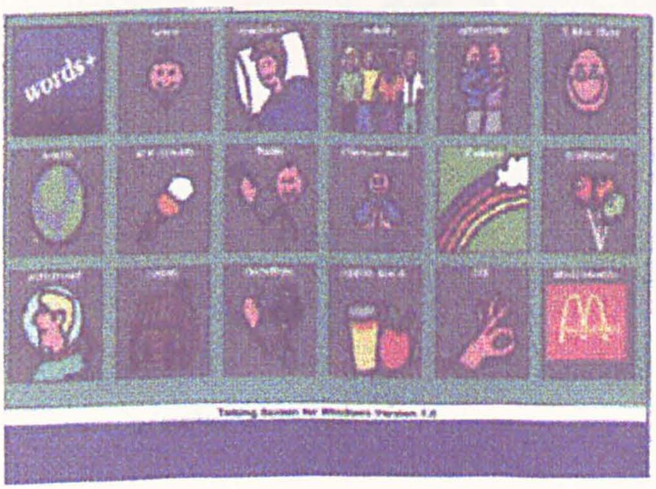
For users who are less able the system can be run in either symbol only mode or in a combination mode, with text and graphics. One *unique* feature of the "DynaVox" is that it offers prediction in not only text mode but also in symbol and combination modes. As the user selects symbols a 'smart palette' of symbols dynamically changes on the basis of the last input. If prediction is successful the user can quickly select a symbol from those suggested as opposed to having to trawl through a huge vocabulary. Thus, interaction is potentially speeded.

The symbol system, "DynaSym" is a rule based system which includes special visual cues for people with memory difficulties. Simple animation of symbols can also be used. In total there are some 2000 symbols within the "DynaSym" vocabulary.

Fig 26c, An example of the DynaSym symbol set

Setting it apart from other commercial systems, "DynaVox" has a rudimentary tracking system called a "Transaction Counter". This simply keeps a tally of the frequency of key presses as a simple tally chart might. Although not approaching the complexity of tracking system, known as the Automated Response Tracking System (ARTS), incorporated within Easy Speaker it does provide information on frequently selected utterances. This might enable areas of weakness to be concentrated upon and strengths to be consolidated.

Fig 3.27a, Talking Screen for Windows

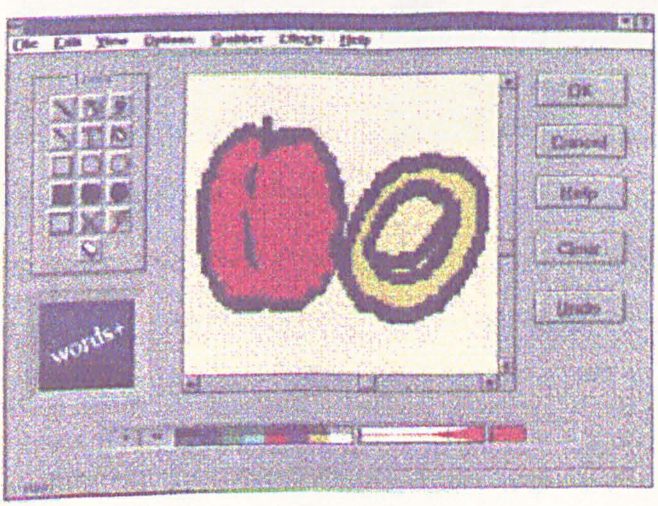


"Talking Screen for Windows" is probably the most advanced commercially available ACD today. The system is very similar to Easy Speaker (1990) and was first developed in 1993 by "Words+ Inc." in that it operates on a similar hypermedia principle. It provides for both digitised and synthesised speech in a similar way to the "Gus!" system. In addition environmental control devices can also be operated.

Fig 3.27b, The maximum 128 symbols per screen

By default the system makes use of "Mayer-Johnson Wordless Edition Book" with some 1600 symbols. The actual symbol system is similar to Rebus bar with 16 block colour graphics. Symbol density can be altered from between 1 and 128 symbols per screen. The a wide range of abilities can be encompassed.

Fig 3.27c, The built-in symbol editor



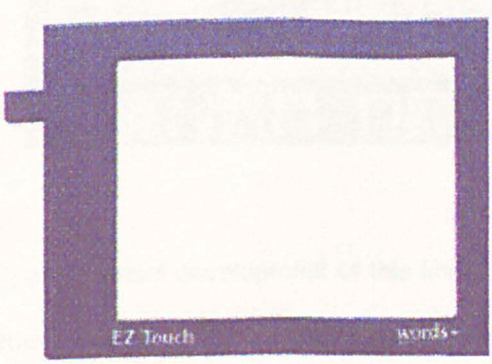
The system also allows for video images to be replayed when associated with a given symbol. The example given is selecting an image of John F. Kennedy and seeing and hearing him deliver an excerpt from his "We choose to go to the moon" speech. The system also has an in-built image editor for creating and modifying simple symbols. On-line help is also provided.

As "Talking Screen" runs under Microsoft Windows (as does Easy Speaker) it can make use of any IBM PC compatible with appropriate sound card and other Windows extensions, e.g. touch screens, alternative input and output devices etc... Words+ solution has been to package the software in to an NEC UltraLite Versa notebook.

Fig 3.27d, The System 2000/Versa

Running on such a powerful but small and highly portable device "Talking Screen" can provide access to a wide range of potential ACD users. With the addition of a touch screen overlay such as the "EZ Touch" the Versa or any portable for that matter becomes a level easier to operate for those with reasonable motor control.

Fig 3.28, The EZ Touch screen overlay for portables



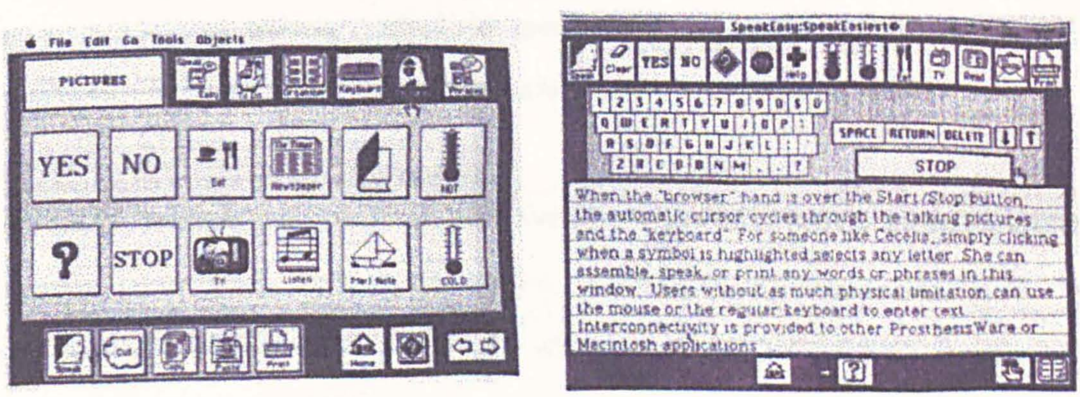
The "EZ Touch" simply clips over the screen of any portable computer and plugs directly into the mouse port in order to mirror the functions of a physical mouse in Microsoft Windows. In addition the unit requires no external power source other than that provided by the mouse port itself.

Advanced Approaches to ACD Implementation

In reviewing a representative sample of commercially available ACD's the generic features of such devices have been highlighted. However, no individual device provides 'all things to all men': there is still room for improvement. Even state-of-the-art devices such as "Gus!", "Cameleon", "Talking Screen" and Easy Speaker cannot encompass every conceivable need. However, there are notable novel and advanced approaches within research and commercial devices which are noteworthy of mention.

As discussed many available ACD's make use of synthesised speech output some systems allow for text to be input and linked with a given symbol. This potentially allows for either a symbol based system or for a text based system. However very few systems offer both. One piece of research "Prosthesisware" that incorporates both interfaces is Chute's (1988) Apple MACintosh based ACD "Speak Easy". This can enable the client to grow with the device, perhaps beginning with symbology and moving on to text as was the original intention with Chute's "Prothesisware" designed for expressive aphasia brought about through head injury.

Fig 3.29, Chute's SpeakEasy offers numerous interfaces



A commercial development of this idea can be seen in the "DynaVox" which can make use of either a textual or symbol based interface, or even a combination of the two.

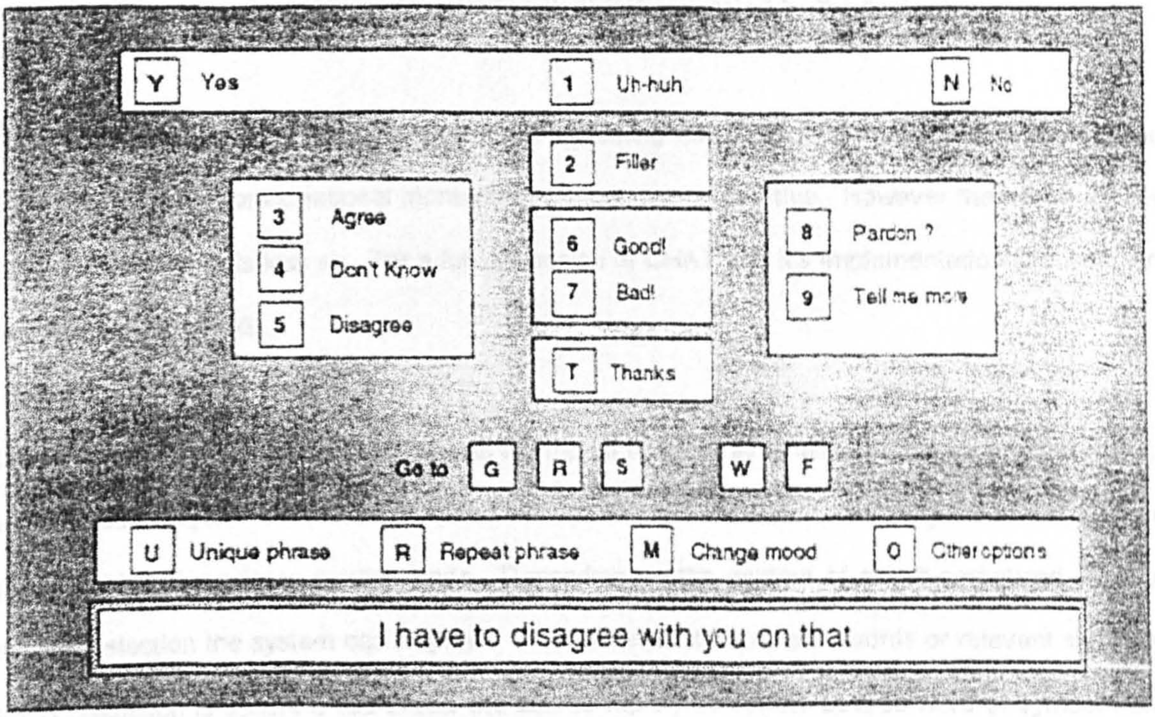
Some of the more advanced text-to-speech devices also allow for tone to be set. For example a button can be pressed which alters the tone that of an "elated speaker". Many such systems also attempt to differentiate between various type of phrase. For example, questions, comments or instructions by altering intonation. One such example is the "Liberator" which allows for many intonnative aspects of speech to be set. Obviously this remains a goal for *all* devices making use of digitised speech. Without a parallel technique the only soluble alternative would be to re-record each utterance with different intonnative qualities.

Another useful innovation is the ability to alter the physical size of the symbols used. For example, "Talking Screen" can re-size symbols so that they take up between 1 and 128 screen locations, or from an A4 sheet per symbol, to postage stamp dimensions respectively. In addition the ability to play other types of multimedia file, such as video images, and not just digitised sound is perhaps the icing on the cake.

However many of these advancements concern the physical aspects of input and output rather than 'true innovation'. One of the few more far reaching advancements has been that made by Alm (1988) in developing CHAT (Conversation Helped by Automatic Talk). CHAT attempts to address the overriding problem of dramatically reduced conversational momentum when using *any* ACD. With natural conversational speeds generally approaching or surpassing 150 words per minute many ACD's are typically credited with speeds or around 15 words per minute (Foulds 1980). It has been found that interactions are likely to fail if rates drop below this boundary.

With CHAT 'smalltalk' and 'fillers' can be automatically generated while the user concentrates on something useful to say. Typically this is used within opening and closing interactions. This enables conversational speeds to be kept high whilst maintaining and strengthening the social aspects of the interaction. The system allows the tone of the interaction to be set and the level of automaticity of 'canned phrases'. That is, phrases can be automatically selected from a set repertoire or conversely they can be constructed by the user albeit at a reduced conversational rate. Certain 'canned phrases' can be associated with a given person and used when talking to them alone. This enables the listeners name for example to be inserted into a 'canned' phrase.

Fig 3.30, How the user sees 'canned' phrases in CHAT



Despite CHAT having been based on conversations modelled with Augmented Transition Networks (ATN) where this theoretically enables the system to predict the next conversational move there are potential arguments that could be levelled with the use of such an automated approach.

By the application of an automated approach one could potentially draw a parallel between some aspects of CHAT and Facilitative Communication.

Felce (1994) summarises FC as :-

> '... [involving] a facilitator supporting an individuals hand, wrist or arm to provide some level of hand-over-arm or hand-on-arm guidance as he or she [the client] spells out words, phrases or sentences by pointing to letters on a letter board or pressing keys on a keyboard.'

One could consider CHAT when operating in a more automated mode and making use of conversational fillers as being a facilitator, albeit a computerised one. Felce (1994) noted that how the majority of interactions mediated though Facilitiative Communication were likely to represent conversation of the facilitator and not the client. In a similar way one might argue that CHAT suffered from similar drawbacks as fillers etc... are in actual fact generated by the computer.

However, the concept of 'canned' phrases is appealing and is an undoubted and much needed way of keeping up conversational momentum without being repetitive. However the notion of fully automatic selection is less so. For a full discussion of CHAT and it's implementation see Alm, Arnott and Newell (1992).

In a similar vain to CHAT, "DynaVox" has attempted to solve essentially the same problem by using prediction algorithms in a bid to speed conversational momentum. Uniquely these can operate in either textual mode or symbol mode. Depending on the context of a half completed word or symbol selection the system can display a 'smart palette' of completed words or relevant symbols. If the prediction is accurate this allows the user to rapidly select the desired word or symbol from those displayed.

However, potentially the same criticism applied to CHAT might be levelled. Namely, that by suggesting what to select next the user may be *led* in to saying something they didn't intend as a result of the ease in with which the offered alternative can be selected. This may be especially true of the group such methodologies might offer most hope to, namely the cognitively impaired.

What is not in dispute is the need to speed communication rates in users of ACD's given the 'compressed' nature of their operation. Systems such as CHAT and prediction schemes such as those used in "DynaVox" are crucial to future development and will ultimately be a part of future ACD's.

Another essential ingredient [in the mind of the author at least] is the necessity that devices should record details of their usage so that a users progress might be monitored. However commercial devices which do this are scarce, and those that do make only a token attempt. The "DynaVox" and its "transaction counter" typify the lack of attention paid to this area. However the "transaction counter" does allow for simple frequency monitoring of key/symbol presses and does suggest that commercial devices will eventually follow this route.

THE NEED FOR FURTHER DEVELOPMENT

With recent and continuing advances in computing power it makes sense to attempt to harness this potential and to produce an ACD capable of storing large vocabularies of high quality speech accessed by dynamic displays of a given symbol set. In terms of digitised speech these might typically represent around three hours worth or more of recording time comprising of over 3500 individual words.

Given such a large high quality vocabulary the next concern is how to represent it to the user in a way that they can easily understand. How can dynamic displays of an alternative symbol set be traversed and how might the symbols themselves be implemented ? Research has suggested that the use of alternative symbol sets are viable, and that the more concrete these symbols the better (Clark 1981). Later research has suggested that these symbols should be of high quality and preferably colour (Mirenda and Locke 1989) and use should be made of symbols and icons already seen in the real word, e.g. brand logos etc... Reichle, Sigafoos and Remmington (1991). Such research ideals have already been implemented in Easy Speaker (1990) with digitised pictures, and to a certain extent in leading-edge commercially available ACD's such as "Talking Screen".
Having chosen a symbol set to represent digitised speech segments the next problem is of providing an interface which provides quick dynamic access to a large symbol set at low cognitive load. Increasing the symbol density or making use of 'shift' like arrangements is not the answer for many users with obvious cognitive deficits. The interface methodology implemented in Easy Speaker is based on a hypermedia metaphor and hopes to provide a low cognitive load way of accessing a large symbol set. In addition such an interface is hoped to provide secondary cognitive training benefits. Again such interfaces are only now beginning to appear in commercial devices costing many thousands.

However despite leading-edge commercially available devices now offering graphically based access to digitised speech vocabularies there is little hard evidence, other than anecdotal, that users benefit markedly from such an approach. This stems from the fact that *no* commercially available devices has any sophisticated tracking that records users interactions with the system, and thus is available for inspection, assessment and monitoring of progress. This might allow for areas of weakness to be concentrated upon and areas of ability strengthened. It also allows for modifications to be implemented that might enable increased performance.

Tracking of users interaction with the system was considered crucial and was built into Easy Speaker from the 'ground-up'. Indeed tacking was considered of as equal importance to high quality photorealistic graphics and digitised speech. Easy Speaker tracks each interaction with the system and these tracks can therefore be used as an assessment tool. Cursor movement is tracked as a measure of efficiency, and therefore to infer learning and motor skill acquisition. Utterance generation speed is also measured and again taken to infer learning and improvements in motor skill. Each utterance that is output is recorded, along with performance measures. The measures Easy Speaker records, by means of the Automated Response Tracking System (ARTS), for each utterance are :-

- Utterance count
- Track of where in the vocabulary hierarchy the user has visited
- The length in words of each section of speech in the utterance produced
- How far the mouse has moved before each graphic and therefore section of speech was selected
- The time when each section of speech was selected (24 hr clock)
- The difference in seconds and milliseconds between each speech section selection
- The cumulative timings for each speech section selection
- The total time taken to select the whole utterance before outputting it as digitised speech
- How far the mouse was moved in total before the utterance was output
- The length in words of the utterance output
- The number of symbols that were selected that represented speech
- The number of symbols that represented links to other parts of the vocabulary hierarchy
- Number of symbol select button presses
- Number of *say* utterance button presses
- Total number of button presses
- How far the mouse was move in order to produce each word on average
- Average number of seconds required to produce each word in the utterance
- The predicted number of words per minute based on the utterance just output

These measures are recorded unobtrusively whilst normal interaction continues and can be analysed at a later date. Indeed with use of standard tasks they formed the main tenet for measurement of 'the usability' of Easy Speaker.

With increasingly powerful generally available hardware the goal then is to get away from 'solid state' devices to ones which are based around software instead, and are as such highly customisable. Later they also have the potential to be upgraded to later and hopefully better versions.

These are the goals which the author has attempted to strive for in the development of Easy Speaker. With new technologies such a system can now be implemented on generally available hand-held systems such as the Fujitsu "Stylistic 500" :-

Fig 3.31, The Fujitsu "Stylistic 500" pen-based mobile computer running Microsoft Windows

With such high ideals one must attempt to justify each of the components that make up a alternative communication system such as Easy Speaker. That is, the symbol set used, the interface methodology and form of speech output. The system must also be usable and produce other measurable tangible benefits for the use.

As regards symbol set common sense is backed up by previous research suggesting that 'obvious', more concrete, symbol sets are those which are best implemented. None can be more concrete that actual photos of things in the real world which for many already have a standard meaning.

SUMMARY

Within this chapter the relative merits of both common forms of ACD output have been considered. Namely synthesised and digitised speech. It has been suggested that digitised speech can offer many advantages over synthesised with the availability of current hardware.

The choice of symbol set is a crucial aspect of design, and with this in mind common symbol based systems have been discussed. The consensus being that the more concrete the system the better, as abstract systems impose a given learning curve which must be climbed. The most concrete system was suggested to be photorealistic representations of the real object, person, or event.

A selection of the best hardware based ACD's available today has been reviewed. These range from the most simple to the most advanced. With the level of advancement typically being mirrored by their intended user populations. In a similar way software based ACD's have also been reviewed.

Advanced approaches in both commercial and research based ACD's have also been highlighted in terms of the innovativeness and relative merits. The need for further development has also been aired, with the empathises being on the actual interface metaphor a new device might use.

This metaphor was suggested to be hypermedia. However the next crucial argument to be answered is whether a hypermedia style interface is the best choice for implementing dynamic displays for cognitively impaired users ? Can a theoretical argument be proposed that might support the use of hypermedia and can this be supported by data from actual users ? It is to these questions that I now turn.

CHAPTER 4: Why the hypermedia metaphor ?

In the previous two chapters I have explored the notion of Learning Disabled Augmentative Communication Device (ACD) users having differing cognitive and adaptive abilities, and reviewed a representative sample of current commercially available computer based ACD's.

However, as highlighted in reviewing many ACD's little account is *seemingly* paid to the differing cognitive and physical abilities of potential users. Most *general* ACD's are *usually* accessible only by the more able due to their inherently complex interface and physical demands placed on users. For example, many ACD's require that users have a relatively good grasp of reading and writing, or that they can make constructive use of a limited, highly abstract symbol set, which they have learnt in order to access a finite underlying vocabulary.

The notion of ACD usage for this particular group is sound, but what is needed is a new interface metaphor which makes no assumptions as to the potential users abilities, and allows easy access through concrete symbolology.

I hope to suggest that this new metaphor should be hypermedia. Although hypermedia style interfaces have been implemented in several commercial devices no theoretical rationale has been forwarded to account for their successes with users. In addition, users have not been evaluated during interactions with such systems in order to provide more objective data. In this chapter I hope to propose why and how a hypermedia interface metaphor should be implemented. Later in chapter nine performance data with Easy Speaker will be examined in order to strengthen theoretical arguments forwarded.

Although a hypermedia metaphor is appealing both in theory and practice with this group, research in office and education based environments has suggested potential drawbacks. These centre around the ease with which unimpaired users can become disorientated, or "lost in hyperspace". Such potential drawbacks will be discussed in relation to LD users of a system such as Easy Speaker.

INTRODUCTION

With the recognition that many previous ACD's have been only usable by the most able, what is needed is a new interface metaphor that would enable many more to make use of their facilities. A low demand, high productivity interface would enable even the most handicapped of the non-vocal Learning Disabled population to use such a device as a communication tool. This new interface is suggested to be provided by a hypermedia style metaphor. However, justification for adopting an apparently complex interface must be outlined from theoretical, practical and experimental standpoints. A brief history of hypermedia, and the cognitive framework that lies behind it will be discussed, as will the relevance of the approach to the LD population. Finally, the "proof of the pudding", will be discussed in relation to evaluating the success of the approach and potential short-comings.

A BRIEF HISTORY OF HYPERMEDIA

Although hypermedia[1] is currently in vogue, with not a computer now apparently being sold without reference to its "stunning multimedia capabilities", the concept itself is a least half a century old. Vannear Bush, president Roosevelt's science advisor in the 1940's, is widely credited with defining the modern concept of hypermedia. Bush homed in on the problem of information-overload within many governmental and industrial documents generated following the end of the second world war. His overriding goal was to make cross-references within documents both easy to create and traverse. The system Bush envisaged was a virtual information system which was built up from an intricate web of connections. These connections would allow the user non-linear access to the information the system contained. Bush regarded these connections as, "an enlarged intimate supplement to memory" (Bush 1945). The hope was that this web of connections would help the user assimilate the information more rapidly due to its increased richness.

The system envisaged, "Memex", was designed to run on microfilm, with connections in the text of each slide referring to text within other slides and so on ... When the user selected a special piece of text on one slide they would be shown related material on another slide. These special pieces of text are now referred to generically as "hotspots". A good example of such a connection might be between an uncommon word on one slide and a description of its meaning on another. Bush thought that selection of these hotspots would best be achieved by utilising eye tracking technology, and hence reduce the effort required on the part of the user. Unfortunately due to technical limitations of the time the system was never implemented.

Having outlined the *birth* of hypermedia, before moving on to discuss systems that achieved reality it is useful to 'flesh out the bones' of Bush's ideals. To help illustrate how such hypermedia systems function at a concrete level it is useful to consider them alongside more traditional information processing strategies.

---

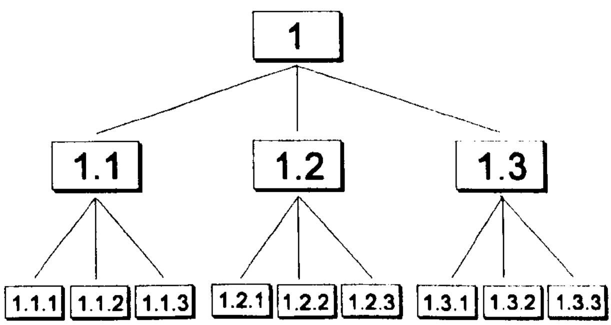
1 Hypermedia is a generic term and is taken to encompass hypertext, and aspects of modern multimedia

The first and most easily understood is the linear information processing model. A good analogy can be drawn between reading a novel, where you start at the beginning and read through to the end in a linear fashion, e.g. Chapter 1, 2, 3, 4 ...

Fig 4.1, A linear information processing model

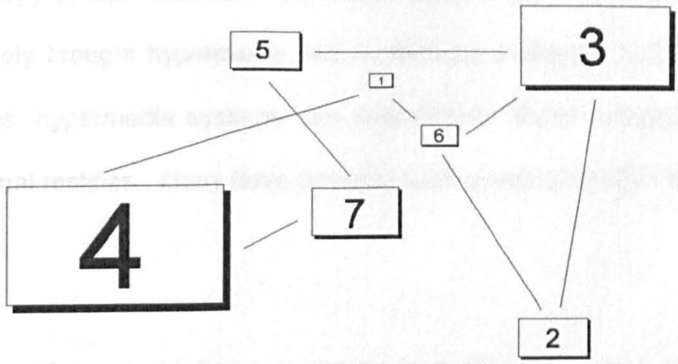
A slightly more complex enhancement is the two dimensional information processing model. This can be compared to a text book which is broken down into ever finer units, e.g. Introduction [1], chapter 1 [1.1], introduction to chapter 1 [1.1.1], body of chapter 1 [1.1.2], summary of chapter 1 [1.1.3] and so on ...

Fig 4.2, A two dimensional information processing model



Finally comes hypermedia which is based on a three dimensional information processing model. Although a first glance a 3D information space can appear complex, in essence it is just as easily understood as the two previous models. Using Bush's concept of microfilm, each number below might represent a separate slide. Slides which contain related information are connected, either directly or indirectly by being connected through other slides. Such connections then form an intricate 3D web which can be traversed in a non-linear manner by activating hotspots on each slide.

Fig 4.3, Hypermedia: A three dimensional information processing model



However, despite being well defined conceptually, hypermedia and its 3D information spaces were not to reach fruition until some 23 years after Bush.

The *reinvention* and promotion of hypermedia was vigorously led by Ted Nelson in during the mid 1960's. Nelson in common with Bush shared a utopian vision for hypermedia, with Nelson coining the term "hypertext". Nelson's over ambitious dreams of systems such as "Xanadu", where the worlds literature would be stored in a massive 3D information space, with an infinite number of connections were never realised.

Encouraged by Nelsons avid promotion Douglas Engelbart produced his "Human Augmentation System" in the later half of the 1960's. This system had many modern hypertext features such as, point-and-click, expanding outlines, multiple windows, remote collaboration, and used a mouse as the preferred control device (Engelbart 1984). During the same period Andries van Dam produced early electronic books which included dynamic graphics and three dimensional animation (Yankelovich et al 1985; van Dam 1988). Restricted by computational power and the inevitable cost of systems at that time, both remained fully functional but experimental, and were never made widely available.

It was not until the mid 1980's when there was a third wave of enthusiasm for hypermedia, championed by Jeff Conklin, that the full potential for hypermedia was realised. This was largely due to the increasing power to decreasing cost ratio of available computer systems capable of using hypermedia software. The starting pistol for hypermedia was fired by Apple Computer when they gave away a free copy of Bill Atkinson's HyperCard system with every system purchased after 1987. This effectively brought hypermedia and multimedia authoring to the masses. With advancing technologies, hypermedia systems can now include digital images, video clips, digital audio, and even virtual realities. Many have deemed such systems to offer true multimedia capabilities.

Given that the ethos of hypermedia has survived for over 50 years, and probably for that matter centuries, what makes the notion of such three dimensional webs so appealing. Can their theoretical basis outweigh their surface attractiveness ?

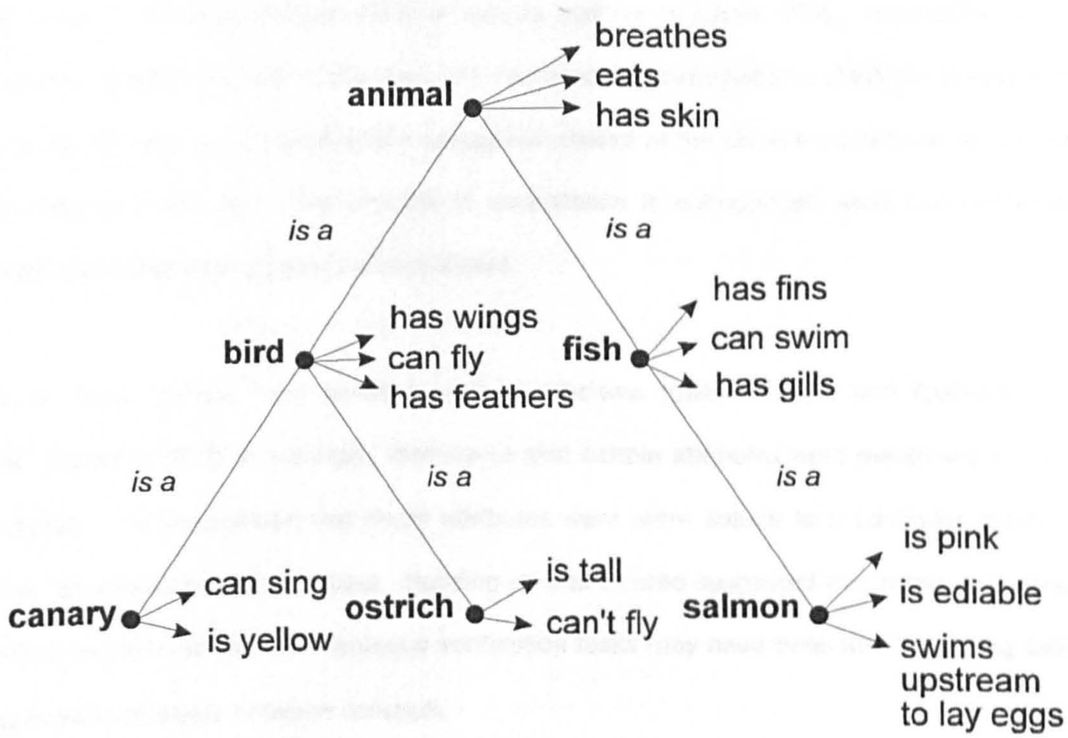
THE COGNITIVE FRAMEWORK BEHIND HYPERMEDIA

Bush, not only outlined the goals for a rudimentary hypermedia system, but also hinted at its theoretical premise. That is, he suggested that such a system might act as "an enlarged intimate supplement to memory". By which he was undoubtedly making reference to the associonist nature of the human mind, as reflected by the title of his article, "As we may think" (Bush 1945).

However the theories to more fully explain these notions only began to emerge some 15 or so years later during the "cognitive revolution" in psychology. Indeed it was during this period that hypermedia reached critical mass; it was intuitively right and it now had cognitive theoreticians who could argue why it had such great potential.

These theories concentrated upon the organisation of knowledge within the human mind. They focused both on how knowledge is represented and how these representations are structured. The majority of these theories were born as the result of Collins and Quillian's (1969) Network theory and model of knowledge representation. Their original model was brilliant in its simplicity as it followed a two dimensional information processing model.

Fig 4.4, A model of the semantic network proposed by Collins and Quillian (1969)



Collins and Quillian suggested that semantic memory was constructed from a set of concepts represented by nodes in a network (indicated by the black dots above). They also suggested that properties for each node were stored at the node level in the most economic way possible (indicated by the arrows above).

Drawing on earlier findings by Bousfield (1953), and Bower et al (1969) in addition to their own on sentence verification tasks Collins and Quillian suggested that people tend to store knowledge in an organised way according to meaningful categories. Bousfield 1953 for example, had suggested that when people are presented with a list of words in no particular order they tend to recall them later by their category. For example, they might recall all the words to do with flowers in one block etc... Bower et al (1969) had suggested if the words were presented in ordered categories originally recall was even better.

Then according to Bush's assumption, "as we may think", if knowledge was indeed stored in a semantic network, a hypermedia system which is also organised as a richly interlinked structure should be synergistic and the information it contains could be both accessed and assimilated more easily. That is, the system should have an almost intuitive structure. If any information within the hypermedia system is novel to the *user*, the framework is potentially in place for a new concept node to be formed, and its appropriate properties created at the correct location on an individuals internalised semantic net. The process of assimilation is unimportant, what is however is the structure the knowledge takes once assimilated.

However, later authors have levelled various criticisms toward Collins and Quillians original model. Conrad (1972) for example, discovered that certain attributes were mentioned more often by subjects. This suggested that some attributes were more salient to a particular person than another, as indeed we might expect. Building on this Conrad suggested that some of Collins and Quillians response latencies in sentence verification tasks may have been due to varying salience as opposed to distance between concepts.

Rosch (1973) asked people to rate the typicality of different members of a concept, and found that some were found to be much more typical than others. For example, a Robin was considered to be a better example of a bird than a Canary etc... It was found that subjects took longer to verify less typical members than typical.

Both of the aforementioned are, logical, and fair criticisms of the original model. However later models have failed to depart significantly from that outlined by Collins and Quillian. Later models built around the original theory in general allow for individual differences in the structure of knowledge. An individuals knowledge may be structured slightly differently at a concrete level, but from a theoretical standpoint may share a common base. Differences between theories are small and many have argued that it is extremely difficult to determine attributes of many concepts anyway (Hampton 1979, McNamara and Sternberg 1983).

Such criticisms are important as they highlight the difficulty in predicting in what manner an individual has either consciously or unconsciously organised their own semantic memory. Thus in attempting to construct a hypermedia system consisting of a 3D information space that is synergistic and widely applicable is just as problematic.

One confounding aspect is that hypermedia systems are three dimensional spaces, and the majority of models of knowledge representation are two dimensional. It is however highly likely that knowledge representation is, as hypermedia is, based upon a 3D information space. Such a system would be likely to be highly efficient in terms of non-redundancy of concepts etc...

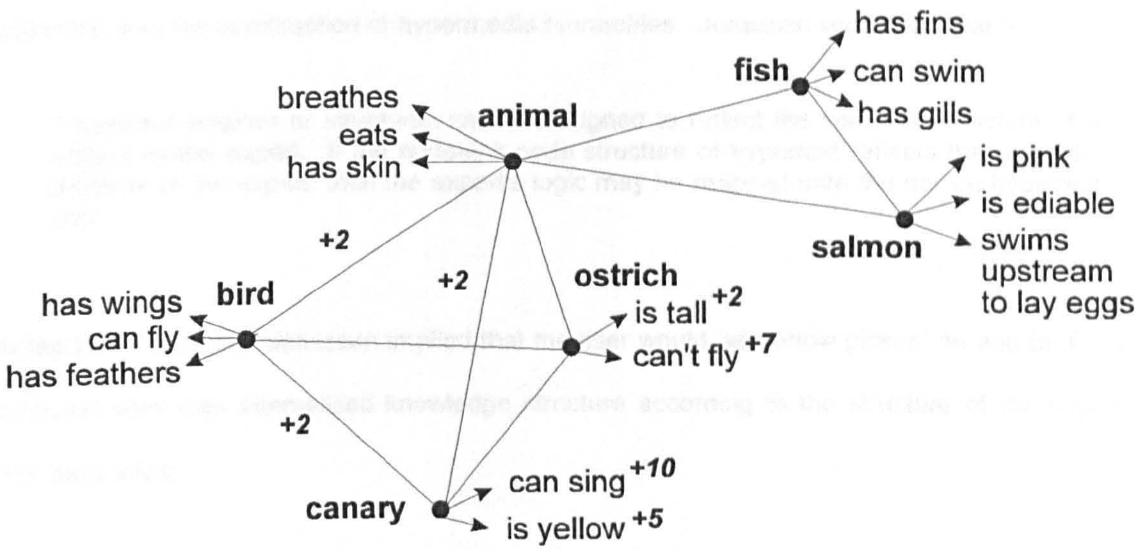
Kintsch (1970) in studying semantic memory hinted at the relationship between semantic networks and 3D hypermedia spaces suggested that :-

> "The meaning of a word is known only if it is contrasted with the meaning of a related or opposed term. For example, colours have meaning in relation not to some physical constant but only to each other as different cultures have different ideas about colour ... The meaning of a term is determined by the way it is stored in memory - as a list of pointers to other terms."

By stressing the importance of pointers, or links, Kintsch outlined how and why a hypermedia system with its intricate links might function. However it was not until the development Notecards (Halasz 1988) that a model such as Kintsch's was applied to a real world hypermedia system. In Notecards a series of cards were linked to form a semantic network, which was covertly imparted to the user as they used the system. Rada and Barlow (1988) building upon this notion consider each card in a HyperCard stack, as a node within a semantic network. They go on to suggest that *x then y* rules can be implemented using structured browsing of decision trees.

Although more specific models have been proposed it is still useful to adapt Collins and Quillians original model due to its generic predictions. For example, based on Collins and Quillians original model a simple 3D representation might look something like :-

Fig 4.5, A hypothetical 3D network hierarchy based on Collins and Quillians original model



As suggested by Conrad (1972) and Hampton (1979), some properties are more salient than others for a given concept. Using such a 3D model I propose I hope to account for why some concepts appear to be more easily accessed than originally suggested by the original model. For example, using the original model to verify that a Canary was an animal this would need to be inferred from the fact that a Canary is a bird, and a bird is an animal. However, using a 3D model there is an alternative root, i.e. the more direct one, a Canary is an animal. To add another level of complexity to the model one might suggest that the paths between concepts might be assigned weightings favouring a particular route. For example, one could verify a Canary was an animal (+2), but one might do so by first deciding it is a bird (+2), and that a bird is an animal (+2). These weightings between paths might be unique to individuals, and in addition subject to change. Analogously an individual may favour a particular route through a hypermedia hierarchy.

Hypermedia then is intrinsically meshed with the concept of semantic networks as a way of representing knowledge, and with semantic memory itself and the way in which we assimilate and retrieve information. Semantic memory according to Tulving (1972) is, "A system for receiving, retaining and transmitting information about meaning of words, concepts and classification of concepts".

Jonassen (1989) outlined the importance of incorporating semantic network models such as those presented, into the construction of hypermedia hierarchies. Jonassen suggested that :-

> "Hypertext engines or structures can be designed to reflect the semantic structure of a subject matter expert. If the node-link-node structure of hypertext reflects the semantic structure of the expert, then the expert's logic may be mapped onto the novice hypertext user".

By the term "mapping" Jonassen implied that the user would 'somehow pickup' on and be likely to formulate their own internalised knowledge structure according to the structure of the hypertext they were using.

Michon (1972) suggested how such representations are related to learning by the assertion that, "Learning is most frequently represented within this approach by means of growing networks, where growing means the addition of new distinctive features at new decision nodes". Michon tested this notion experimentally by presenting eight pictures to children and adults to see how they grouped them, classified them, and what relationships they constructed. Adults tended to follow the accepted taxonomy, whereas children grouped pictures according to relationships that were relevant to them. Such differences in individuals conceptual frameworks highlights how material within any hypermedia hierarchy needs to be structured according to the end users level cognitive model of the domain. For example :-

Fig 4.6a, A 32 year olds model of understanding of given categories based on Michon's assessment tasks

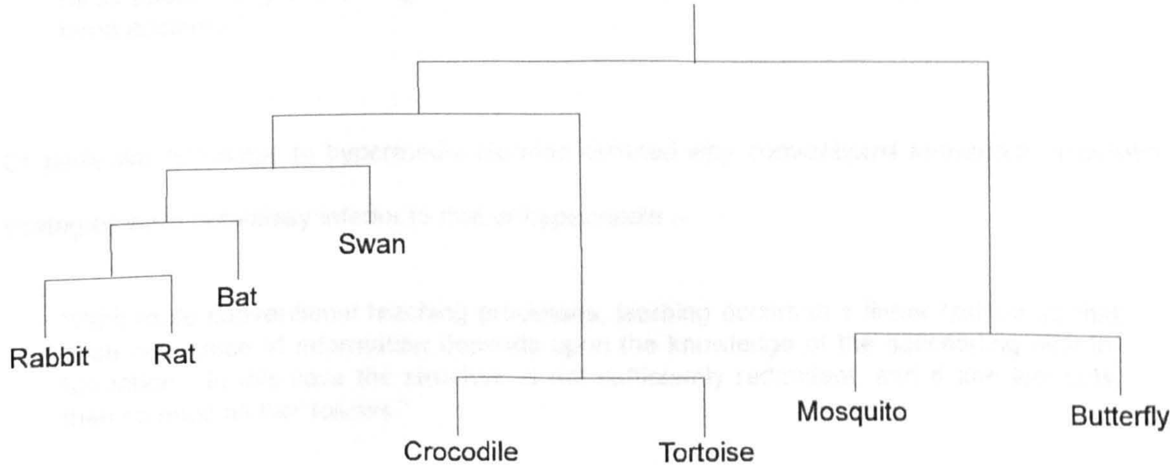
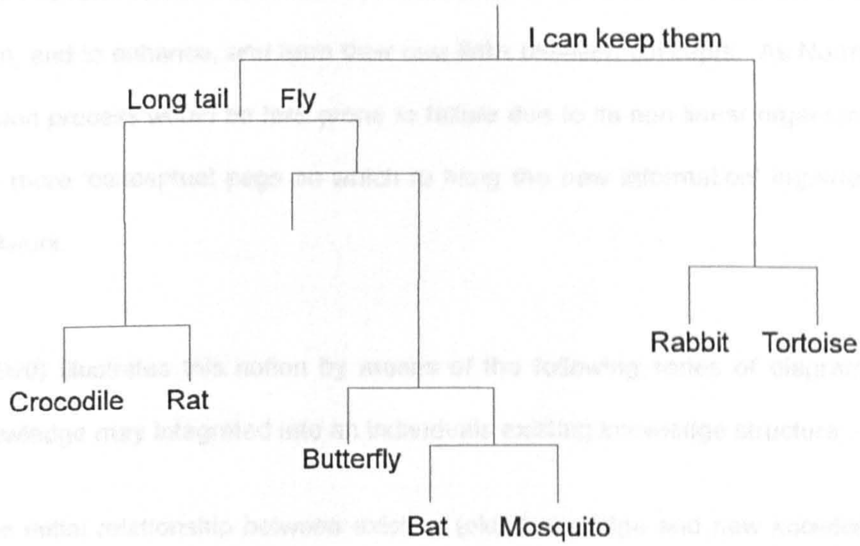


Fig 4.6b, A 7 year olds model



By utilising such cognitive mapping techniques Michon clearly indicated the importance of paying heed to the persons current level of understanding and to the way in which they structure their personal knowledge base. Applying the model one can also see how such a semantic network is in a constant state of flux as new concepts are added and integrated, or linked with other existing knowledge.

Norman (1975) operationalised the process of integrating new knowledge into an existing network :-

> "... learning should add the new material by attaching a conceptual framework within memory to the old previously acquired material. That framework should then be elaborated upon, filling in the rough web with finer structures until a complete network has been acquired."

Of particular relevance to hypermedia Norman outlined why conventional knowledge acquisition strategies were potentially inferior to that of hypermedia :-

> "With more conventional teaching processes, learning occurs in a linear fashion so that each new piece of information depends upon the knowledge of the succeeding new information. In this case the structure is not sufficiently redundant, and if one link fails, then so must all that follows."

Norman stressed the importance of organisation of information, and implied that browsing though a hypermedia network would enable the individual to take on the links in the knowledge provided by the system, and to enhance, and form their own links between concepts. As Norman points out this assimilation process would be less prone to failure due to its non-linear organisation. In short the user has more 'conceptual pegs on which to hang the new information' imparted via the hypermedia network.

Churcher (1990) illustrates this notion by means of the following series of diagrams suggesting how new knowledge may integrated into an individuals existing knowledge structure :-

Fig 4.7a, The initial relationship between existing (old) knowledge and new knowledge to be acquired

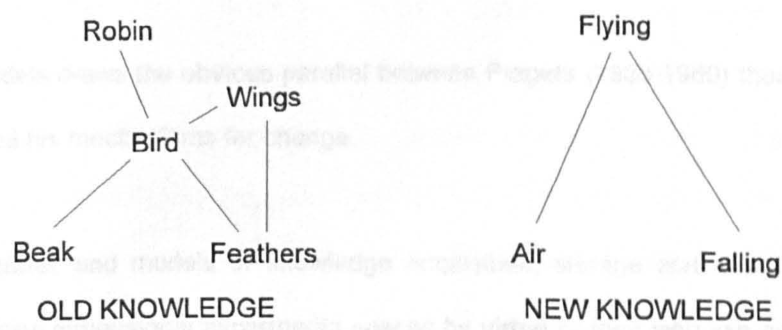


Fig 4.7b, The initial integration, or linking of new knowledge to existing concepts

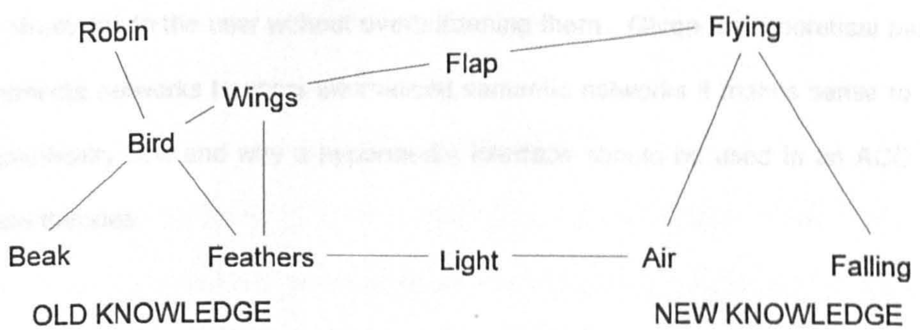
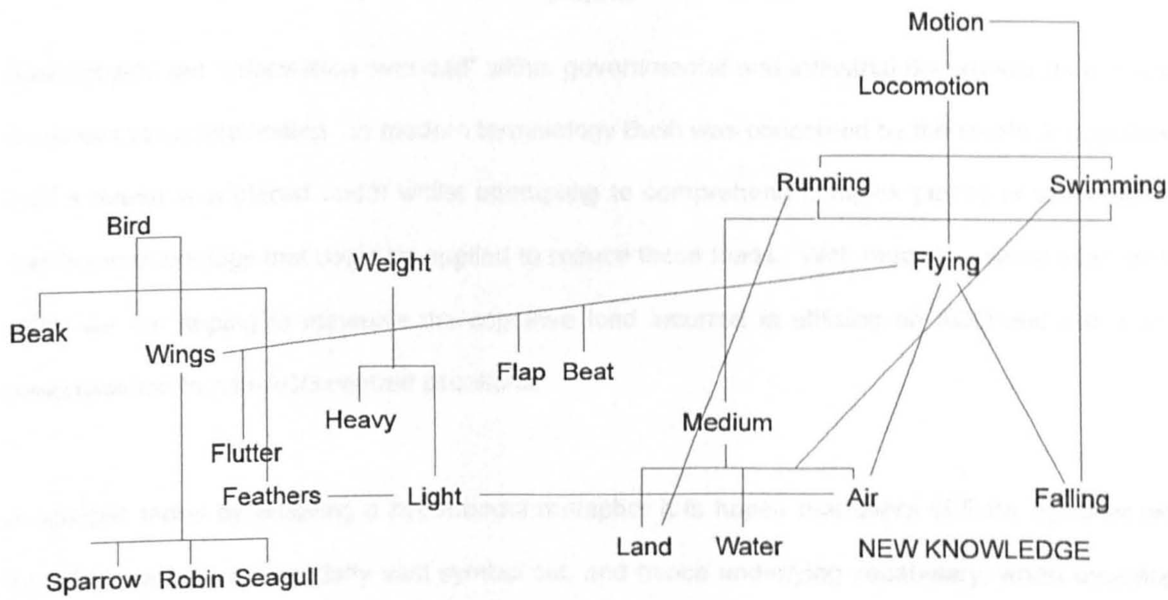


Fig 4.7c, The full integration of new knowledge through elaboration



Use of such models draws the obvious parallel between Piagets (1886-1980) theory of intellectual development and his mechanisms for change.

Given such theories and models of knowledge acquisition, storage and manipulation one can postulate that three dimensional hypermedia spaces by virtue of their web like organisation might provide a synergistic way of organising and accessing large symbol based vocabularies. In addition it might also provide an almost covert, or incidental, way of imparting new concepts and knowledge structures to the user without overburdening them. Given the theoretical base for alikening hypermedia networks to users internalised semantic networks it makes sense to attempt to describe specifically how and why a hypermedia interface should be used in an ACD with reference to these theories.

How and Why Hypermedia Should be Used in an ACD for Learning Disabled Users

Bush singled out "information-overload" within governmental and industrial documents as a major constraint to understanding. In modern terminology Bush was concerned by the levels of cognitive load a reader was placed under whilst attempting to comprehend complex pieces of information, and the methodology that could be applied to reduce these loads. With much the same overriding ethos we are hoping to minimise the cognitive load incurred in utilising an ACD and maximise generalisation to non-ACD centred situations.

In specific terms by adapting a hypermedia metaphor it is hoped that users of Easy Speaker will be able to access a potentially vast symbol set, and hence underlying vocabulary, whilst incurring a relatively low cognitive load. Rather than presenting a user with say 128 postage stamp size symbols it is far better to reduce the number of symbols shown at any one time and to group those that are shown in to meaningful categories. Special symbols shown within a given category might indicate a link, or relationship to other concepts, and when clicked on might transport the user to that category and so on... Thus a three dimensional information space is born, with 'hotspots' providing the means of navigation between concepts. As suggested by Bousfield (1953) and Bower et al (1969) these categories of symbols should be organised in to meaningful groupings. For example, symbols concerning "drinks" might be shown together on one screen. Symbols related to drinks might be also be shown acting as links to related categories, e.g. things you put in drinks such as "ice cubes", "sugar", "milk" etc...

The dilemma occurs however when choosing which symbols to include in categories, with some authors indicating that constructing a widely applicable category is near impossible (Hampton 1979 and McNamara and Sternberg 1983). Michon (1972) highlights the differences between individuals conceptual organisation and how this is subject to change over time. Thus suggesting that even within accepted categories individual variation must be accounted for.

Jonnasen (1989) however suggested that hierarchies should be, initially at least, organised to reflect the subject knowledge of an expert. The hope being in the case of a cognitively impaired user an *idealised* organisation might be mapped on to the users own. Theoretically this might allow for rapid navigation of the hierarchy and for improvements in the users own knowledge organisation. Despite the difficulty of constructing categories of symbols, research has suggested that it is possible to identify certain categories or symbols that normally go together (Conrad 1972 and Rosch 1973). However, within these categories some links might be more salient to the user than others (Conrad 1972 and Hampton 1979) as a result of previous experience.

In addition to grouping symbols according to *accepted* categories it is also important to ensure that the links between categories are as rich as can be achieved with respect to the users ability. Norman (1975) suggests that the richer the links between pieces information the more likely we are to learn it. He paraphrased this as being like, 'Having more pegs on which to hang information'. Norman points out how traditional learning tends to occur in a linear fashion, where what we learn next is dependant upon what we already know. If a link fails lower down the chain the chances of learning newer related information successfully is reduced. By utilising a three dimensional space with an enriched number of links between related concepts there is always an alternative route to a given piece of information. Indeed Kintch (1970) stressed the importance of such relationships and put them on equal standing with the concepts themselves.

In addition to enabling synergistic navigation of a large three dimensional space in order to select symbols and eventually output speech, the secondary hope for such systems is to create or strengthen areas of the individuals knowledge structure. That is, off-task, to strengthen correct associations between categories and concepts and ultimately to integrate new knowledge. With Churcher (1990) suggesting that assimilating new knowledge can be accomplished by forming many and varied links with existing knowledge structures. Rada and Barlow (1988) go so far as to suggest a close relationship between semantic networks and three dimensional hypermedia stacks. In fact they consider each card within a HyperCard stack as a node being within a semantic network.

Many, especially in the field of Artificial Intelligence would argue that not only is knowledge structured and accessed according to a semantic net, that it is from this network that language is produced and understood. Some authors argue that transformational grammar itself can be based around knowledge represented as a semantic network, e.g. Katz and Fodor (1963). Norman and Rumelhart (1975) have constructed semantic networks that break down sentences into concepts and the relationships between them. By such representations of concepts they have suggested how language understanding can be modelled based on a semantic network.

Authors such as Shank (1975) and Rumelhart and McClelland (1986) have created computer models which make use of semantic networks to understand language. That is, they extract meaning from language by analysing concepts and the relationships between them. Some would even go so far as to suggest that all understanding can be based around semantic networks (Johnson-Laird 1993).

Undoubtedly there is an inextricable link between semantic networks, knowledge representation and understanding and language reproduction and understanding. Hypermedia allows a semantic network to be hidden from an ACD user but still offer all the theoretical benefits. Namely ease of navigation of concepts and properties in order to construct meaningful utterances symbolically which can then be output as high quality speech. Despite many potential users severe cognitive and linguistic handicaps there is no reason to assume they do not make use of internalised semantic networks to extract meaning from the world around them. By use of such a representation it is hoped that not only will one device production be boosted but also that users' understanding, or internalised semantic representation, will also be enhanced.

As modern computers enable instantly identifiable photorealistic images to be used as symbols representing categories and properties this undoubtedly adds to the ease with which the hierarchy can be navigated. By selection of symbols the hierarchy can be traversed, utterances built up, and finally these can be generated as digitised speech.

The underlying rationale for Easy Speaker in relation to the hypermedia metaphor is :-

❶ To select concepts so as to match the current understanding of the user.

❷ To correctly organise the concepts and therefore impose an experts structure that will hopefully be transferred to the user.

❸ That once structured to match the user, they should be able to navigate the hierarchy intuitively. That is, their internal knowledge representation should be a close approximation of the hypermedia hierarchy and vice-versa. That is, both should build upon the other.

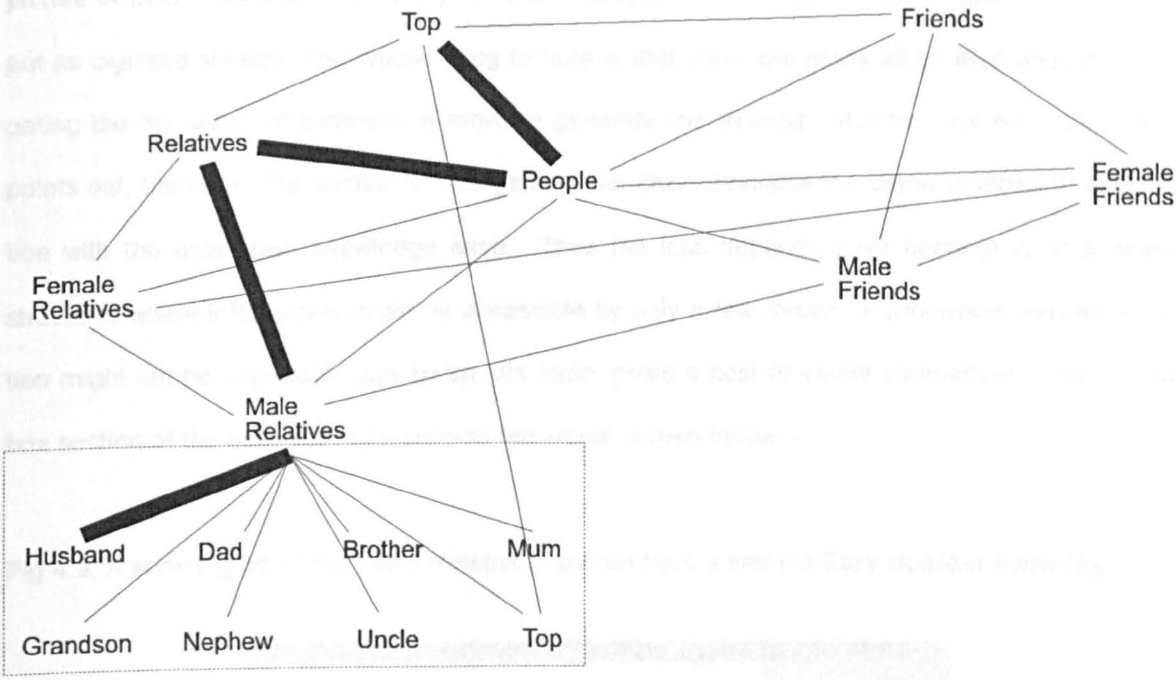
❹ New concepts and properties should be able to be gradually introduced and incorporated into the Easy Speaker hierarchy and hopefully therefore the users own knowledge hierarchy.

❺ Because the hierarchy is intuitive, it should be easily traversed by the user for a low cognitive effort. Furthermore because relevant photorealistic images are used to represent concepts, links between concepts and properties, i.e. digitised speech, the user should have enough cognitive resources remaining to devote to navigation, planning and to absorbing the structure and phonetic qualities of the digitised output etc... That is, when a resource based cognitive model is applied.

❻ To gradually increase the number of *correct* links between concepts in the hope that these would be absorbed by the user. An increased number of links would not just enrich the knowledge base but also allow for alternative navigational strategies.

By applying such a rationale one might construct a large three dimensional hypermedia space tailored to a specific individual. In order to clarify how such a space might be constructed and navigated a small section relating to “Male Relatives” from an actual hierarchy is shown below :-

Fig 4.8, A concept map of an actual Easy Speaker hierarchy relating to "Male Relatives"

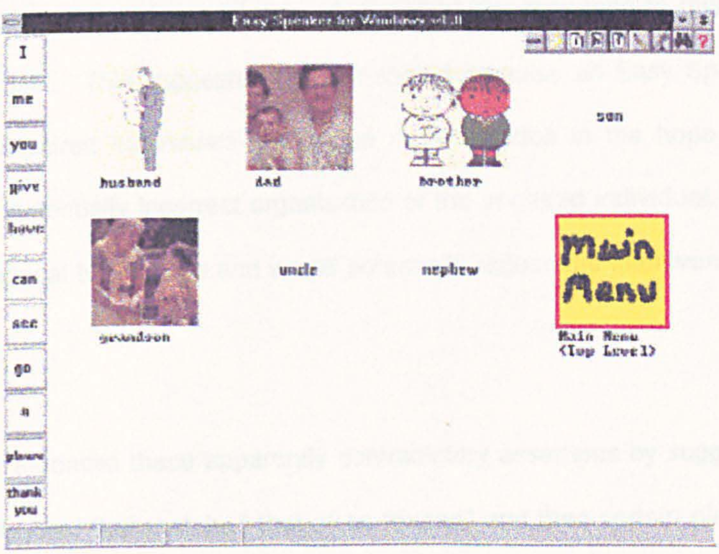


In the concept map above one can see the three dimensional relationship, or links, between concepts concerning male relatives. Each of these concepts would represent a separate screen within Easy Speaker, and the links between different concepts would be indicated by a symbol surrounded by a red box on each of these screens. The user would simply click on this symbol to activate the link and be transported to that concept screen and so on ... Thus the hierarchy could be traversed.

Above, under the concept "Male Relatives" appears a number of properties associated with that concept, i.e. "Husband", "Dad" etc... These are represented within Easy Speaker by symbols of the properties they represent, e.g. "Husband" is a picture of the users husband etc... Each one of these pictures is married with a piece of pre-recorded digitised speech mirroring the actual picture. The user can select these pictures building up a visual representation of the utterance they wish to generate. When this is complete they can output the utterance as digitised speech. That is, the pieces of digitised speech associated with each of the individual pictures.

In the example above, in order to *say* "Husband" the user might choose to navigate to the concept of "Male relatives" by clicking on the links, "People", "Relatives", "Male relatives" as indicated by the heavy black links. Once at the concept screen for “Male relatives” they may then click on the picture of their “Husband”, a property of “Male relatives”. The word “Husband” could then be output as digitised speech. The crucial thing to note is that there are many alternative ways of navigating the 3D space of concepts in order to generate the desired outcome. As Norman (1975) points out, the larger the number of navigational solutions available the better in terms of integration with the users own knowledge base. Thus the less dependent we become upon a linear structure, where information might be accessible by only a few routes. Furthermore new information might not be accessible only by an ‘old route’ given a host of viable alternatives. The dotted box section of the map above represents the actual screen below :-

Fig 4.9, A screen grab of the “Male Relatives” screen from a sample Easy Speaker hierarchy



As Michon (1972) concluded when studying the organisation of concepts and the relationships between them in adults and children the concepts themselves and the links between them were unique to the individual. That is, we impose our own structure on knowledge. In terms of Easy Speaker this would suggest that it is crucial to customise any structure used to the individual concerned in order to maintain its intuitiveness, and reduce the cognitive load in traversing the hierarchy. As is the case with the above screen the organisation of concepts has been tailored to a particular individuals current level of understanding of those concepts and their relationship to other related concepts.

However organising any hypermedia hierarchy to suit a specific user is not simply a question of logistics, but rather one of the possible implication this may have for reinforcing an incorrect understanding of given concepts. In addition one might inadvertently add to these anomalies by increasing the number of other related concepts associated with the base concept which itself is incorrect. What is needed is a balance between the users current understanding and the introduction of new concepts and modification, or removal of incorrect ones. That is, in terms of the associations between concepts and the concepts themselves. For example, on a screen that concerned pets, you might not want a picture of a car to appear, even though the pet may have at one time or another travelled in the car.

In relation to Michon's conclusion Jonassen (1989) suggested that if the hypermedia structure was constructed according to the representation of an expert this may lead to this organisation being mapped onto the user. This suggested that we should organise an Easy Speaker hierarchy according to an unimpaired individuals knowledge representation in the hope that this would be mapped on to the potentially incorrect organisation of the impaired individual. However, logically this view is paradoxical to Michons and would potentially reduce the intuitiveness of the hierarchy for the individual.

One can however reconcile these apparently contradictory assertions by suggesting that an initial structure be constructed that matched that of an "expert" and then certain pieces modified to suit individual users. By use of the term "expert" one refers to a *normal* language user who is ideally matched with the user in terms of age, gender, social class etc... The hope being to maintain the overall structure but simply to incorporate extra links between concepts that were specific to the user. The overriding goal would be to reduce the number of these incorrect links and incorrect concept organisations.

Indeed this approach was applied in Easy Speaker intervention throughout. Once organised the hierarchy should reflect the users internal knowledge representation and vice-versa. The long term goal was to gradually introduce new concepts through Easy Speaker at their correct point within the hierarchy and to enrich the number of links between those new and existing concepts.

The use of photorealistic images and digitised speech was intended to help make the hierarchy more memorable to the user (Mirrenda and Locke 1989). The use of images allowed for given concepts on the hierarchy to be represented pictorially, as opposed to representations that are of a more abstract nature such a BLISS or Rebus etc... These images were also where possible relevant to the user (Reichle, Sigafoos and Remmington 1991). For example, the concept of pets may have been signified by a scanned picture of the users pet cat. When clicked on the may have led to a screen representing properties related to the users pets, e.g. a picture of the user feeding the cat etc... As these properties were clicked on these images would be appended to the utterance under construction and then output as digitised speech.

If need be the user could navigate between screens, selecting concepts and properties from various locations within the hierarchy. That is, they could use an alternative path or strategy for achieving the same end, and would therefore be less constrained by their lack of knowledge of a certain set of concepts that followed a certain path through the hierarchy. Indeed this is one of the many benefits of implementing a 3D information space as Norman (1975) points out. Given the ability to select different paths through the hierarchy to achieve the same end a users knowledge of related concepts and properties would be likely to be reinforced. Simply put, the 'web' would be continually 'woven' throughout interaction as a user would be less likely to fail in achieving a given goal. "Success leads to success", and success is one of the strongest reinforcers available.

The use of digitised speech further improves how memorable the hierarchy is. For example, a high quality piece of speech associated with a particular high quality image is likely to associate the two, together with its geographical location within the overall hierarchy. For example, an image of a cat, triggering a piece of speech saying cat followed by a purring sound is likely to be remembered very vividly.

As the organisation of *any* Easy Speaker hierarchy attempts to closely mirror the archetypal three dimensional information spaces outlined, one might conclude that it represents a use useful and relevant interface. Furthermore, the availability of such a system might allow users easy access to such a large knowledge base despite their inherent handicap. With the system being constructed to be both easy to use, visually appealing and audiologically compelling it is hoped that such a multi-sensory approach would further build upon the potential benefits offered by implementation of a hypermedia metaphor. The hope being that users would not simply use Easy Speaker as an ACD but also to improve their own language and cognitive abilities also.

ADVANCED FEATURES OFFERED BY HYPERMEDIA

Not only can a hypermedia system offer photorealisitic graphics and CD quality digitised speech and sound it can also provide full multimedia if so enabled by the developer. This means that multimedia features such as animated graphics, video clips and music files can be incorporated. Such features would be likely to further enhance the salience of symbols and the system as a whole.

A still photorealistic symbol, for example, may instead of being linked with high quality speech may be linked with a video clip.. Thus, when selected it may play the video and accompanying sound track. Through such advanced features learning would undoubtedly be enhanced and motivation would be likely to be raised. In a similar way animation's could also be used.

By making use of video clips and animation's this also raises the possibility of making use of 'live' symbols versus those of the static variety. The possibility for enhancing cognition and learning is intriguing. The implication for such features in the ACD arena remains to be investigated fully. Their relevance however is undoubted.

In addition to providing for sound and motion, multimedia also offers the ability to play MIDI based music. That is, true, loss less digitally based music that can recreated at the touch of a button. This would enable MIDI music files to be associated with symbols so that the score is played when a given symbol is selected. Potentially this would enable music collections to be built up that could be replayed for the users own pleasure and where appropriate to be shared with others.

Although such enhancements are not encompassed within the software used in the current study they are planned for later versions and further studies. Hopefully a version with such enhancements will be developed so that its potential might be outlined with regard to future directions.

EVALUATING THE SUCCESS OF EASY SPEAKER AND ITS HYPERMEDIA METAPHOR

In order to assess whether real users could operate the system successfully, and therefore infer that a hypermedia interface was almost transparent in use, a range of set tasks was designed and implemented so as to form an assessment tool. These measure various performance indicators, using the in-built tracking system, over a series of sessions so that improvements might be plotted. Indications of on-task improvements could be taken to indicate the successful application of an hypermedia interface. For a full discussion of the tasks themselves see chapter eight and for the results users achieved see chapter nine.

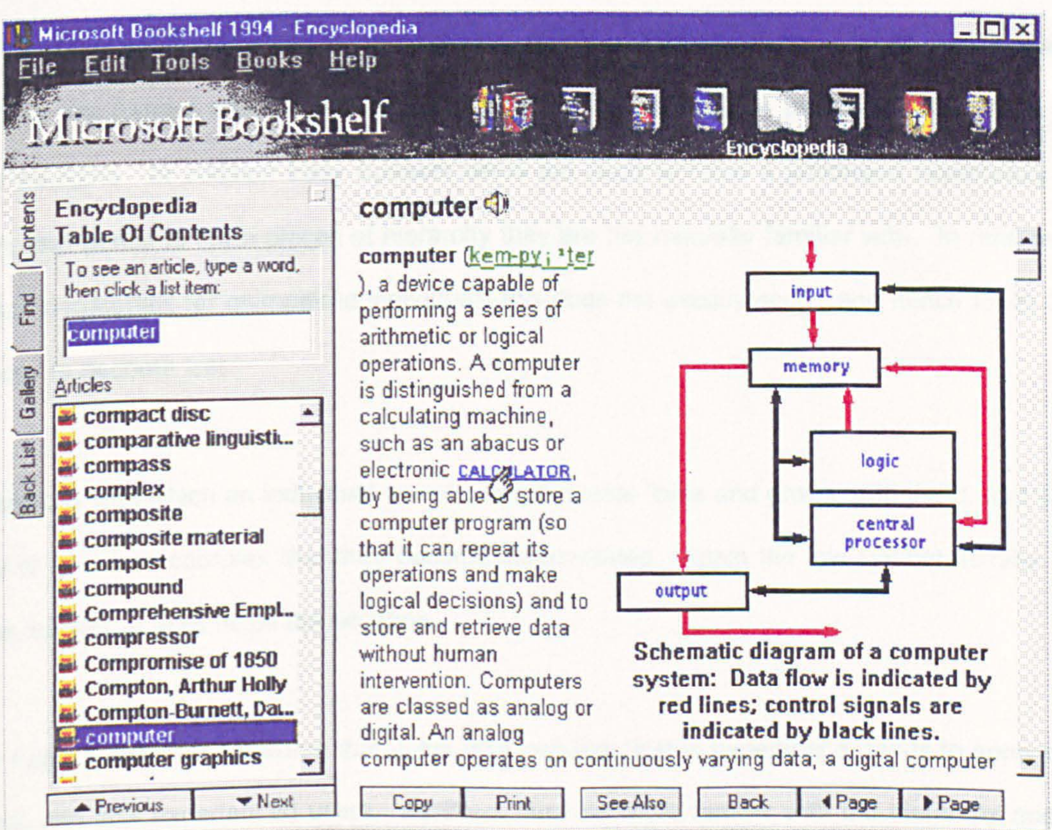
In order to evaluate off-task improvements in language and cognition a longer six month study was conducted with pre and post intervention questionnaires being administered to two of each users *immediate* caregivers. These subjectively questioned caregivers about areas such as expressive and receptive language, attention and memory, and motivation etc... An improvement over the baseline taken at pre-intervention would suggest that prolonged Easy Speaker usage would be likely to have a positive impact both on and off-task. Within the sample group even a small improvement would be personally meaningful to both themselves, their caregivers and, not withstanding the author. Improvements would suggest that a hypermedia style interface can be of some benefit to clients. For a full discussion of the questionnaire design again see chapter eight, and for an in-depth analysis of the results obtained from users see chapter ten.

POTENTIAL SHORTCOMINGS

Despite the enthusiasm for the application of hypermedia and the cognitive framework that can be applied to it, there are potential shortcomings. These have been highlighted in commercial and education based environments where hypermedia systems have been used for information retrieval or teaching purposes. Concerns focus on information overload and anxiety when using such systems (Wurman 1989), and the proneness to become "lost in hyperspace" (Conklin 1987).

Conventional hypertext systems for unimpaired users make use of a number of links within sections of text and graphics. These links, or hotspots, generally transport the user to related material or give a pop-up definition to clarify the subject matter. In essence they are similar to the links within an Easy Speaker hierarchy. For example, Microsoft Bookshelf shown below exemplifies how hotspots are typically indicated, with the word "CALCULATOR" shown in uppercase blue underlined text. To obtain relevant information on the calculator all the user need do is click on the hotspot.

Fig 4.10, A common implementation of hypermedia hotspots in Microsoft Bookshelf



Problems occur when users click on sequences of such links, or hotspots, and as a result become lost within the overall structure. They typically become disorientated and overloaded with the volume of relevant, or irrelevant, information they are presented with. Elm and Woods (1985) describe this situation as :-

> "The user not having a clear conception of the relationships within the system, or knowing his present location in the system relative to the display structure, and finding it difficult to decide where to look next within the system."

Potentially such a situation could occur with users of Easy Speaker, or any other ACD based upon a similar metaphor. The impact of becoming "lost" is perhaps more worrying considering the typical cognitive deficits suffered by LD individuals. Elm and Woods define "lost" as :-

- Not knowing where to go next
- Knowing where to go but not knowing how to get there
- Not knowing where they were in the overall structure of the document

In terms of Easy Speaker the larger the hierarchy, the higher the symbol density, and the richer the links between the more likely an unskilled novice user is to become lost.

However, unlike systems for which the "lost in hyperspace" assertion was coined, Easy Speaker hierarchies are formed with the intention of being relevant to a given user and therefore being easy to navigate. In addition Easy Speaker users are likely to have a prolonged opportunity for practice, and hence to learn pieces of hierarchy they are not naturally familiar with. In relation to educational materials for unimpaired individuals this does not usually occur, and hence leads to a proneness to become lost.

The hierarchy with which an individual uses to communicate 'lives and grows' with them, and so is unlikely to prove so complex that they become disorientated. Again the low symbol density and variable number of links helps the situation.

As one might expect the assertion that users may become "lost in hyperspace" tends to apply only to novice and less experienced users. As users become more familar with the hierarchy and as more idiosyncratic material and links are added they become progressively less likely to become lost. Field trials suggest this to be the case regardless of their IQ, or age comparison scores. For the most impaired users, overall hierarchy size may be reduced, symbol density decreased, and the richness of links watered down. If need be a link back to the highest level can be included on each screen as a further safety measure. Thus for users the possibility of becoming lost can be traded off against functionality.

SUMMARY

In reviewing many previous commercially based ACD's in chapter three the argument against usability over complexity of design was raised. This led to the suggestion that a new interface metaphor might be more appropriate for those with a LD who might benefit from ACD usage. This new interface metaphor was suggested to be hypermedia.

Within this chapter the birth and rationale for hypermedia has been outlined. In support of the applicability of hypermedia a cognitive framework for its success has been proposed. This has been related to why such a metaphor should be used in an ACD for LD users.

Potential shortcomings have been outlined, as has the methodology for assessing the success of an ACD using such an interface metaphor. What remains is to discuss formally the processes through which such a device is developed. In this respect the development processes Easy Speaker went through will be discussed in chapters five and six.

CHAPTER 5: The development of Easy Speaker: A HCI perspective

The characteristics of the user population Easy Speaker was designed for has been outlined, and the relevance of their abilities and weaknesses stressed. Previous commercial and research based ACD's have been reviewed, as have the AAC strategies they employ. A new interface metaphor in the form of hypermedia has been proposed from both theoretical and practical standpoints.

What remains however is to describe how Easy Speaker was developed, and the evolutionary processes it went through. That is, the actual design and implementation of the software shell which allows hypermedia style vocabularies of digitised speech to be built, and represented using photorealistic images as symbols. The relevance of the design process to potential users' and maintainers will be stressed.

Brief account is also paid the hardware requirements needed to utilise Easy Speaker to its full potential.

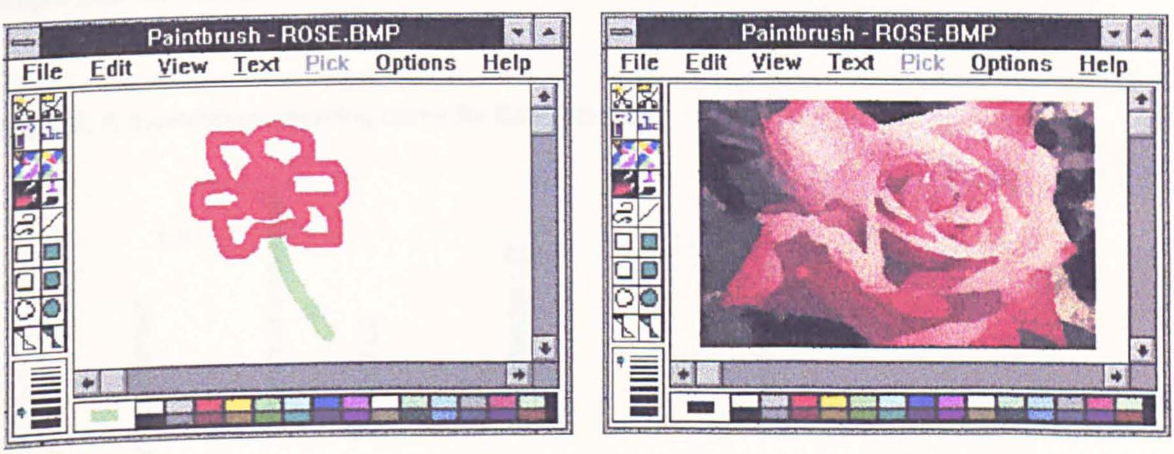
INTRODUCTION

The characteristics of the user population Easy Speaker was intended for has been outlined. Research has suggested that over 90% of the LD population have some degree of communication disorders, whether they be severe or slight. Thus the need for an ACD for this group has clearly been established. However, in addition to communication disorders many suffer from deficits in cognition. Such deficits can include poor short term memory, weak sequencing skills, perceptual problems, poor co-ordination, short attention spans and general slowness of thought. Additionally low motivation and secondary behavioural problems such as challenging behaviours are common.

Overriding the issue of producing an ACD capable of augmenting or replacing an individuals residual speech is the need to take account of these *obvious* handicaps during the design process. As highlighted in reviewing many previous commercial and research based ACD's, designers have constructed devices that can *only* be used by the most able who represent a small fraction of the population who might benefit. As Newell and Alm (1994) point out, "[often] designers have assumed that their users will have the characteristics of a 25 year-old male who has a PhD in computer science and is obsessed with technological gadgets rather than getting on with the job. Systems are a triumph of functionality over usability."

The goal then is to design for those who are least able but potentially have the most to gain from a well designed device. It is a question of good design rather than a compromising design. HCI research has transformed text based office software into easy to use graphical based equivalents over the last few years, e.g. MS Windows and OS/2. Just as in the commercial world the focus has shifted from functionality to usability, so too should it in the ACD arena. Good design ensures that whatever an individuals ability, they should be able to utilise a piece of software at some level. Obviously more able individuals may be able to exploit the software to its full potential. In doing so they should not be hampered by an apparently under powered user interface, in fact the interface should be almost transparent to them. To draw a parallel, even severely learning disabled users can draw a simple picture using MS Paintbrush, however more able artistic users may be able to create a masterpiece :-

Fig 5.1, A well designed piece of software can be used by anyone, regardless of ability



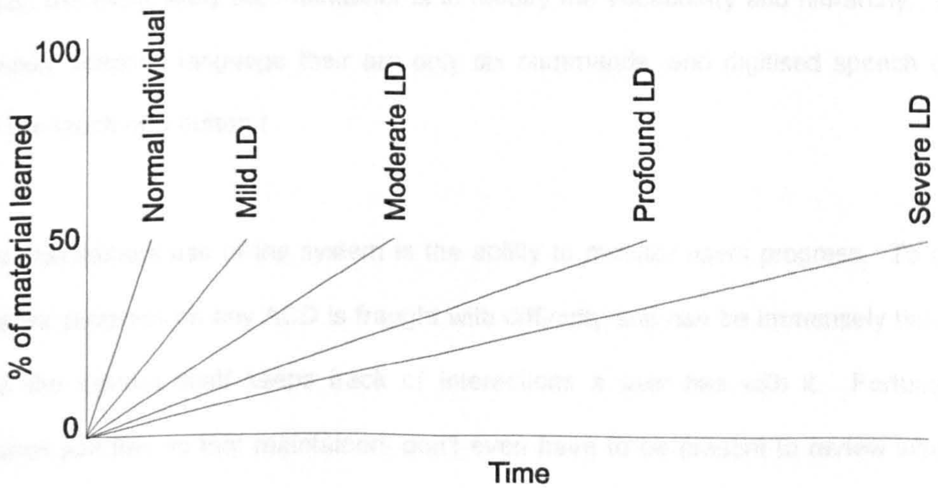
In a similar way a well designed piece of ACD software should be able to be utilised productively by a LD individual despite their *obvious* cognitive deficits.

In order to make a symbol based ACD system accessible it should make use of an accepted symbol set which is easily understood through previous learning experiences. Numerous pieces of research have suggested that these symbols should be photorealistic images of the real object, person or event in question. These symbols are associated with a piece of digitised speech which is *spoken* when images are selected and output triggered. Thus potentially the user is under little load when trying to recall the speech fragment associated with a given picture, as it is already personally relevant to them.

The next question is how to organise such symbols to allow easy, low demand access to a potentially huge vocabulary ? The author has proposed, and discussed in depth the need for a *new* access metaphor. This access metaphor is hypermedia. Again the organisation of the symbols should be synergistic with the users current level of understanding. Thus individuals should easily be able to select the desired symbols from the hierarchy by traversing it. Finally they should be able to output the symbols selected as digitised speech.

If any of the symbols or sections of the hierarchies organisation need be learnt by rote, the learning curve they follow should be favourable even for those with a severe LD. A hypothetical cure might look something like :-

Fig 5.2, A hypothetical learning curve for Easy Speaker



Due to deficits in memory and other cognitive functions, one might suggest that the more LD an individual the longer it will take them to reach a given competence. However, as the interface should be so synergistic in use the percentage learned should only relate to the size of vocabulary they can manipulate, and not to operating system demands per se.

The needs of potential users are clear; ease of use and access to a potentially huge vocabulary. However, within the design process those who maintain the system should also be catered for. It is no use designing an end-user system which is highly usable, if the maintenance of such a system is so tedious that facilitators, caregivers, or health care professionals simply do not maintain system to fit a given individual user, or users needs. For example, to add a new symbol many commercial systems requires that a symbol be hand drawn using a bitmap editing package. This obviously takes time, and in a time pressured service environment such as the NHS this is unlikely to be done. However, if an photographic image can be scanned and digitised in seconds, many more symbols can be added in the time taken to hand draw one. The result in the latter case will be of much higher quality, and is likely to be immediately recognised by the user if correctly selected.

The ease with which digitised speech or sound can be recorded and linked with any symbol is paramount. In common with many advanced ACD's Easy Speaker uses a simple scripting language to construct screens of symbols and the links between them. The simpler the scripting language again the more likely the maintainer is to modify the vocabulary and hierarchy. Within the Easy Speaker scripting language their are only six commands, and digitised speech can be recorded at the touch of a button !

Related to maintainers use of the system is the ability to monitor users progress. To objectively monitor users' progress on any ACD is fraught with difficulty and can be immensely time consuming unless the system itself keeps track of interactions a user has with it. Fortunately Easy Speaker does just this so that maintainers don't even have to be present to review what the user has done !

Having outlined the general software design goals it is useful to move on to highlight recognised user interface design strategies in relation to the evolution of Easy Speaker from both a users and maintainers perspective.

SPECIFIC DESIGN GUIDELINES FOR POTENTIAL LD USERS

Within the field of HCI there are many theories which attempt to formalise what *good* interface design is, some read almost like a recipes for success. However, as one might expect one theory cannot cover all eventualities, so the choice of a correct design model is crucial. Two basic models were used as guidelines during the development of the Easy Speaker user interface.

The first was the Syntactic Semantic Object-Action (SSOA) model when applied to direct manipulation (Shneiderman 1983) of icons on screen to initiate actions. In the case of Easy Speaker direct manipulation concerns the selection of symbols that represent digitised speech segments, and to the navigation of a large three dimensional hierarchy, or vocabulary, made up of pages of such symbols and the links between them.

The second concerns semiotics, or the study of signs and symbols. Semiotics is not concerned with how symbols are manipulated, but with how they are represented. In this case how they are represented on screen to the user. It is the *maintainers* job to select symbols that are understood and synergistic to the users current level of understanding. Arnheim (1972) has suggested that such symbols can represent a visual language and visual way of thinking.

The SSOA model in relation to the design of Easy Speaker and direct manipulation

The SSOA model was originally developed by Shneiderman (1980) to describe the process of programming, and to separate the syntactic aspects from the semantic. When applied to direct manipulation the model proposes that the user has semantic knowledge about what they want to do, and syntactic knowledge regarding how this might be done. To take a concrete example, an individual may want to type short letter on a word processor, and be fully aware of its content (semantic). However, they may not be able to type it as they don't know the method for operating the specific word processor (syntactic). So too in Easy Speaker the user may know what they want to say, but not how to say it using the interface provided. The ultimate goal is to bridge this gap simply.

Syntactically, when using any piece of software the user must maintain and manipulate a myriad of device dependant details, e.g. which mouse button to click on where, which key to press etc... In general syntactic knowledge is varied, device dependent, acquired by rote memorisation, and easily forgotten (Shneiderman 1992). Often commercial ACD systems impose that complex strings of key presses be learnt by repeated paired-associate learning. The "Liberator", reviewed in chapter 3 is a classic example of complex syntactic design, as on occasion up to 10 key presses may be required to output a single word. As Shneiderman notes, "Rote memorisation requires repeated rehearsals to reach competence, and retention over time is poor unless the knowledge is applied frequently." Although Shneiderman is referring to unimpaired users, the implication for complex systems for LD users is even more relevant. The Keep It Short and Simple (KISS) rule should be applied where possible. Thus, anybody, regardless of ability should be able to use the system productively. In terms of operating Easy Speaker all the user need do is move a chosen screen pointer over the symbol they wish to select with the mouse. Click the left mouse button to select the symbol and the underlying digitised speech it represents, or click the right button to output the utterance they have constructed graphically as digitised speech. Obviously the input device need not be a mouse, it could be a joystick or touch screen. What is important is that there is little syntactic knowledge that need be learnt thus increasing the chance that even the least able will be able to consistently operate the system productively. Simply, the means of operating the system should not be so complex, and hence cognitive demanding, as to get in the way of what a user wants to do or say.

Semantic knowledge on the other hand is made up of two components, namely computer concepts and task concepts. For example, a computer concept might include the knowledge that the computer might store data. In the case of Easy Speaker the user will have a good idea that the system might somewhere within the hierarchy, or vocabulary, hold a representation of the phrase they want to *say*. This is a high level concept. Analogous to means-ends problem solving such concepts can be decomposed into smaller units. In Easy Speaker, a user may have a clear idea of what they want to do, and as a result break the concept down into manageable parts. For example, I want to say "Hello Dad", therefore I need to go to the greetings screen, select "Hello", and then to the relatives screen, select "Dad", and so on ... These concepts should hopefully be as natural to the user as to the "experts" who designed the hierarchy in question.

At a lower level are the task concepts. For example, the user may have planned what to do at a higher level, i.e. set the goal, but now has to carry the actions out through task concepts. Using the previous example of generating "Hello Dad", they might have to :-

❶ Move to and left click on the link to greetings symbol
❷ Move to and left click on the "Hello" symbol
❸ Move to and left click on the link to relatives symbol
❹ Move to and left click on the "Dad" symbol
❺ Right click to produce *speech* output

Thus, one can appreciate the need for some degree of task analysis, or at least task understanding. The less steps needed to complete a given task the better, especially for the LD user. In the design of Easy Speaker the goal has been to reduce the need for syntactic knowledge and computer concepts. This has been accomplished by using a visual representation of task objects which can be utilised by direct manipulation. In other words, simple controls and a relevant photorealistic image symbol set. The key is to "Know the user" as Hansen (1971) rightly empathises.

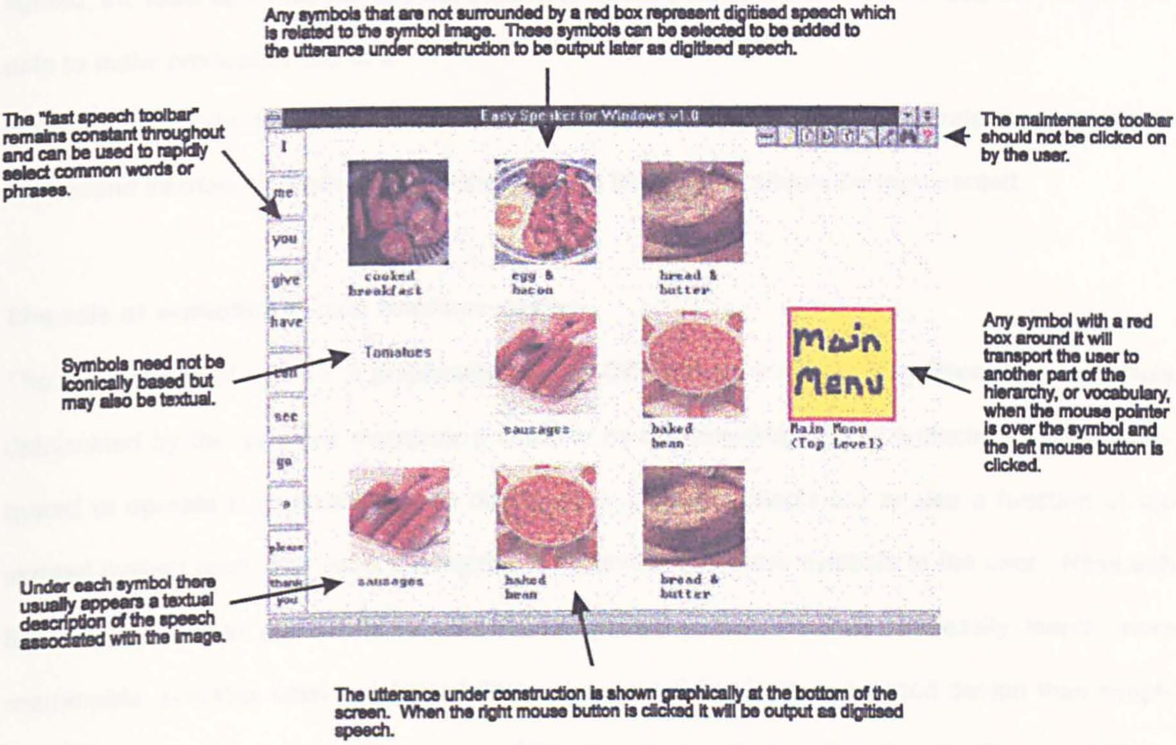
Using the SSOA model Shneiderman (1992) describes seven beneficial attributes such systems might possess :-

- Novices can learn basic functionality quickly, usually through a demonstration by a more experienced user
- Experts can work rapidly to carry out a wide range of tasks, even defining new functions and features
- Knowledgeable intermittent users can retain operational concepts
- Error messages are rarely needed
- Users can immediately see if their actions are furthering their goals, and, if the actions are counterproductive, they can simply change the direction of their activity
- Users experience less anxiety because the system is comprehensible and because actions can be reversed so easily
- Users gain confidence and mastery because they are the initiators of action, they feel in control, and the system responses are predictable

Due to its simplicity and direct manipulation interface Easy Speaker has the potential to provide many of these benefits for those who might appreciate them most, namely LD users.

Users must only learn a small amount of syntactic information in order to operate the interface successfully. These can be best illustrated by the following diagram of the actual Easy Speaker interface :-

Fig 5.3, Syntactic knowledge required on the part of the user



More formally only four essential pieces of syntactic knowledge required to operate Easy Speaker productively :-

- Any symbols that are not surrounded by a red box represent digitised speech which is related to the symbol image. These symbols can be selected to be added to the utterance under construction to be output later as digitised speech.
- Any symbol with a red box around it will transport the user to another part of the hierarchy, or vocabulary when the mouse pointer is over the symbol and the left mouse button is clicked.
- The "fast speech toolbar" remains constant throughout and can be used to rapidly select common words or phrases.
- The utterance under construction is shown graphically at the bottom of the screen. When the right mouse button is clicked it will be output as digitised speech.

The overriding ethos is one of simplicity through direct manipulation of easily understandable, memorable and relevant symbols. Combined with a hypermedia metaphor, with links on screen represented by red bounding boxes, more symbols are accessible than can be comfortably displayed on one screen. By virtue of such organisation and dynamic displays even potential users with marked cognitive and other handicaps should be able to make use of the system. As highlighted, the least able may not use the system to its full potential, but none the less they should be able to make productive use of it.

Having outlined the fundamentals of the syntactic knowledge required to operate Easy Speaker, the second interface ingredient is how the symbols themselves should be represented.

The role of semiotics in user interface design

The functionality of a piece of graphically based ACD software such as Easy Speaker is not solely determined by the interface metaphor it uses, or by the simplicity of the syntactic knowledge required to operate it productively. As discussed in previous chapters it is also a function of the symbol system used, and more specifically the relevance of those symbols to the user. Research has suggested that photorealistic symbols are probably best, as they are easily learnt, more memorable, and less open to interpretation. However, there is more to good design than simply digitising photo's and placing collections of these on screen to be clicked on by the user. Questions such as, how many should be on screen simultaneously, how should links between screens be represented, what colour should the background be, should graphic symbols be accompanied by text, must be resolved.

Fortunately design guidelines for icons do exist and can be applied to graphically based systems such as Easy Speaker. Shneiderman (1992) considers seven features that might be beneficial :-

- Represent the object or action in a familiar and recognisable manner
- Limit the number of different icons
- Make the icon stand out from the background
- Consider three-dimensional icons; they are eye-catching, but also can be distracting
- Ensure that a single selected icon is clearly visible when surrounded by unselected icons
- Make each icon distinctive from every other icon
- Ensure the harmoniousness of each icon as a member of a family of icons

The first recommendation is the cornerstone of Easy Speaker. That is, the use of symbols which are *already understood* by the user as a result of previous learning experiences. The symbols themselves are highly recognisable as they are of around passport photo size and quality and are facsimiles of the actual object, person or event. Logically one would suppose the load they impose in recognition and manipulation would be minimal.

Having championed the use of real images the next question is how many to display simultaneously ? Some commercial systems range from anywhere between four and 128 symbols simultaneously. Obviously the more symbols the higher the demands placed on the user in terms of memory, attention, and general cognitive overhead. This is regardless of symbol set chosen, and is undoubtedly compounded by systems that use abstract symbology. Logically then one would suggest that the fewer symbols were displayed within reason, the better. Given the available screen resolution available, 640x480 pixels in 16.7 million colours, standard symbols were deemed to be 96x96 pixels in size. The size of the standard symbol was chosen so that on screen it was approximately the same size and quality as a passport photograph. Informal observations had suggested that real passport photo's were easily recognised and manipulated correctly by potential Easy Speaker users regardless of their level of handicap.

This meant that eight passport size photo's per screen could be displayed, with these acting as representations for digitised speech or links to other such screens. These eight symbols were variable as each screen dynamically changed as users navigated the hierarchy via the link symbols. However eight high quality symbols was considered to be too few, and may make the user through necessity navigate the hierarchy needlessly looking for a speech symbol that wasn't displayed on a relevant screen due to size/density limitations. The answer was to create a "fast speech toolbar" which contained smaller symbols which remained constant regardless of which dynamic screen the user was on. These 11 symbols were 32x32 pixels and again could be in up to 16.7 million colours. This toolbar was located on the left hand edge of the screen and was easily accessible at all times. The ideal behind the tool was that *maintainers* chose the 11 most frequent utterances that needed to be conveyed by a given user and represented these graphically on the toolbar.

The next issue was one of representing the utterance the user had constructed up to a given point, even though they could choose individual symbols from any available screen and not just the one they were currently on. It was decided that only four symbols would be shown from the utterance under construction, again of passport photo size. The utterance under construction would be shown at the bottom of the screen and would remain constant until output. The fact that only four symbols could be displayed was not considered a limitation for the following reasons :-

- A small number of constructed symbols would allow users to review what they were about to say easily, and would be unlikely to place them under great cognitive load. Options to allow scrolling through longer utterances were considered but discounted due to increased load.
- Symbols need not represent one digitised word, they might represent part of a phrase, a whole sentence or even a short monologue. Thus the fact there might be only four symbols displayed may be unrelated to the number of words *spoken*.
- If need be the user could continue to select additional symbols over the four displayed with the proviso that these would not be displayed in the utterance under construction.

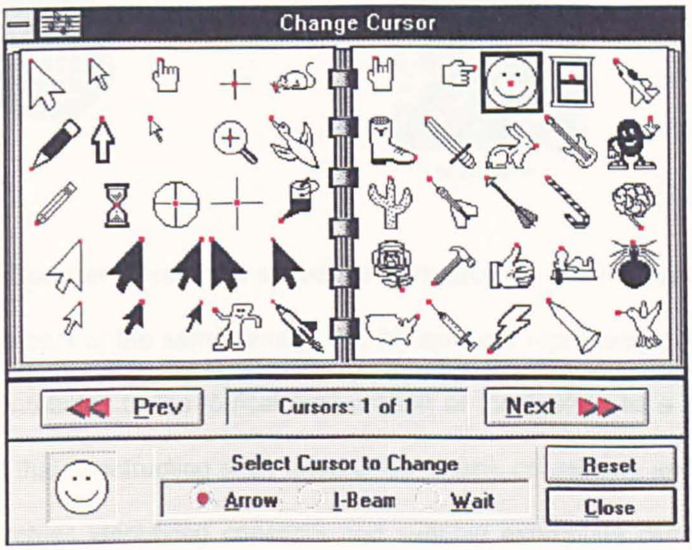
These three choices of symbol size and density are highlighted below :-

Fig 5.4, Choice of symbol size and density

To make each symbol stand out from the background a plain white backdrop was chosen. This would provide high contrast between the majority of the symbols and the background. Symbols that were particularly appealing were those that had depth. For example, those that contained a photo of the individual with a static background, e.g. a picture of them on a beach representing the phrase "I went on holiday to ..."

Ensuring that a single selected icon is clearly visible when surrounded by unselected symbols was crucial. Firstly, symbols that act as links were surrounded by a red bounding box, that was five pixels in width. The hue of the red selected was such that it was almost luminescent, and so could not easily be mistaken for an integral part of the symbol it signified was a link. Any symbols that were not surrounded by a bounding box were signified as being speech symbols. Secondly, various on-screen pointers could be selected, and if necessary these could be user modified. Thus the standard arrow style MS Windows pointer could be replaced by a larger more visible one which could be up to 32x32 pixels in size. This provided a highly visible pointer which clearly indicated which symbol was highlighted and would be selected when clicked on.

Fig 5.5, An example of a Windows utility to change the standard mouse cursor

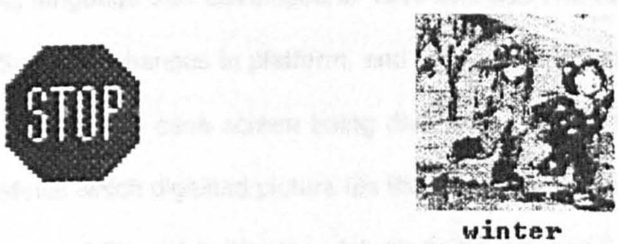


Thirdly the final requirement was that the utterance under construction was clearly visible when surrounded by unselected symbols was accomplished by the fact that its position in relation to other symbols was fixed. In addition users were taught that the four bottom symbols represented graphically the utterance that was about to be output.

The fact that symbols were digitised from photographs in the main ensured that each symbol was distinctive from others. Theoretically this reduced the burden on users to differentiate between alike symbols. For example, additive systems such as Bliss make use of symbols which are very subtlety different graphically, but have a totally different and some times opposing meaning. Thus the load during scanning is potentially increased.

In addition to the content of each symbol differing text under each symbol also reflects the content. For those users who have a basic grasp of written English this is a useful addition, and for those who do not it helps to pair the written text with the symbol. Norman (1991) amongst others has found that text either incorporated within an icon, or under it is more effective than either in isolation in computing environments. Indeed this method is followed in *all* WIMP environments, e.g. MS Windows etc...

Fig 5.6, Text displayed within an icon and text underneath a symbol in Easy Speaker



By their design Easy Speaker hierarchies should be harmonious. That is, symbols on any given screen should be members of the same family, e.g. be symbols representing "cold drinks" for example. Again this contributes to the logical organisation of the system as a whole. However, it must be remembered that constructing such hierarchies is time consuming and incredibly difficult. As in semantic hierarchies identifying concepts and suitable examplars can be difficult and requires careful thought and planning on the part of the designer.

Having discussed both the syntactic knowledge required to operate the Easy Speaker, and the representation of symbols from a user perspective it is important to discuss design issues from a *maintainers* viewpoint.

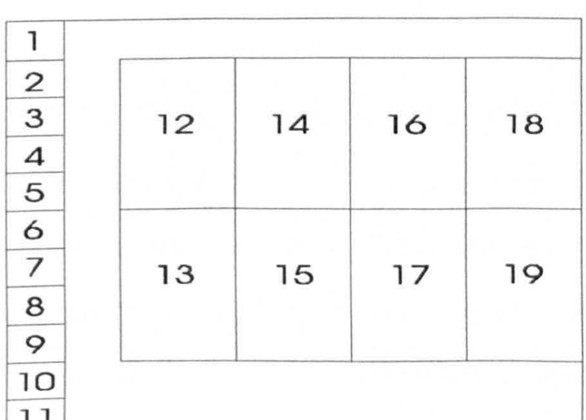
SPECIFIC DESIGN GUIDELINES FOR MAINTAINERS

Prior to the design of Easy Speaker the author examined the practicalities of using existing multi-media authoring packages to implement hypermedia stacks, that used photorealistic images linked with digitised speech. Such packages considered included Guide, Toolbook, Hypercard etc... However many of these packages required maintainers to learn complex scripts and most were unusably slow and cumbersome in use due to their interpreted nature.

This prompted the author to write Easy Speaker in C and C++ to serve as a hypermedia shell with its own simple scripting language. This allowed for very quick user response times even on mediocre equipment and gave total control over implementation of the scripting command set and user monitoring. As Easy Speaker implemented its own unique scripting language it needed to be simple in order for maintainers to learn it, and more importantly to maintain hierarchies for specific users.

The actual scripting language was developed in 1989 and has remained constant despite various versions of Easy Speaker, changes in platform, and changes in choice of compiler. The scripting language itself is based upon each screen being divided up in to a series of pigeon holes. The maintainer then defines which digitised picture file they want to appear in which pigeon hole. Each screen has an ASCII text file which lists the pictures to be displayed on that screen and in which location.

Fig 5.7, The pigeon holes into which pictures can be placed on each Easy Speaker screen



An excerpt from a screen layout, or blackboard, file might look something like :-

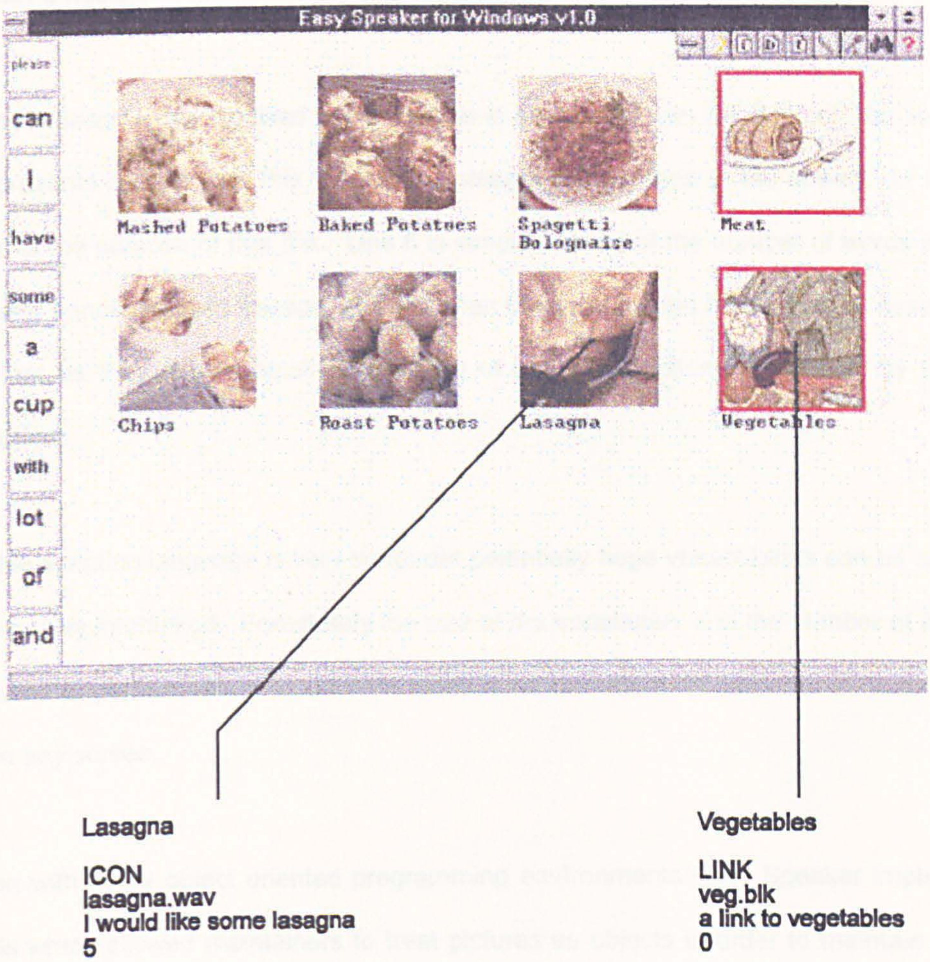
Fig 5.8, An excerpt from the potatoes and pasta screen layout file, "potapast.blk"

```
please.bmp
1
can.bmp
2
i.bmp
3
.
.
spagbol.bmp
16
lasan.bmp
17
```

The screen layout file simply defines which pictures are to be used and in which location they are to be displayed and nothing more. Each picture must have an ASCII text information file which describes whether the picture represents speech or is a link to another screen. When a given screen is activated via the user clicking on a link picture the screen layout file for that link is loaded and the pictures displayed in their appropriate locations. Each information file is loaded for each picture and stored in memory. Then when any picture is clicked on the program can take the appropriate action.

Using the example of the potatoes and pasta screen layout above one can see each information file for two pictures, one representing speech an another a link to another related screen.

Fig 5.9, The information files for "lasagne" and "vegetables"



The information files for all symbols are simple and only contain six lines. The two shown above are "lasan.inf" on the left, and "vegetabl.inf" on the right.

The first two lines of each file serve the same function and contain the text shown under the picture. If there is no second line it is left blank.

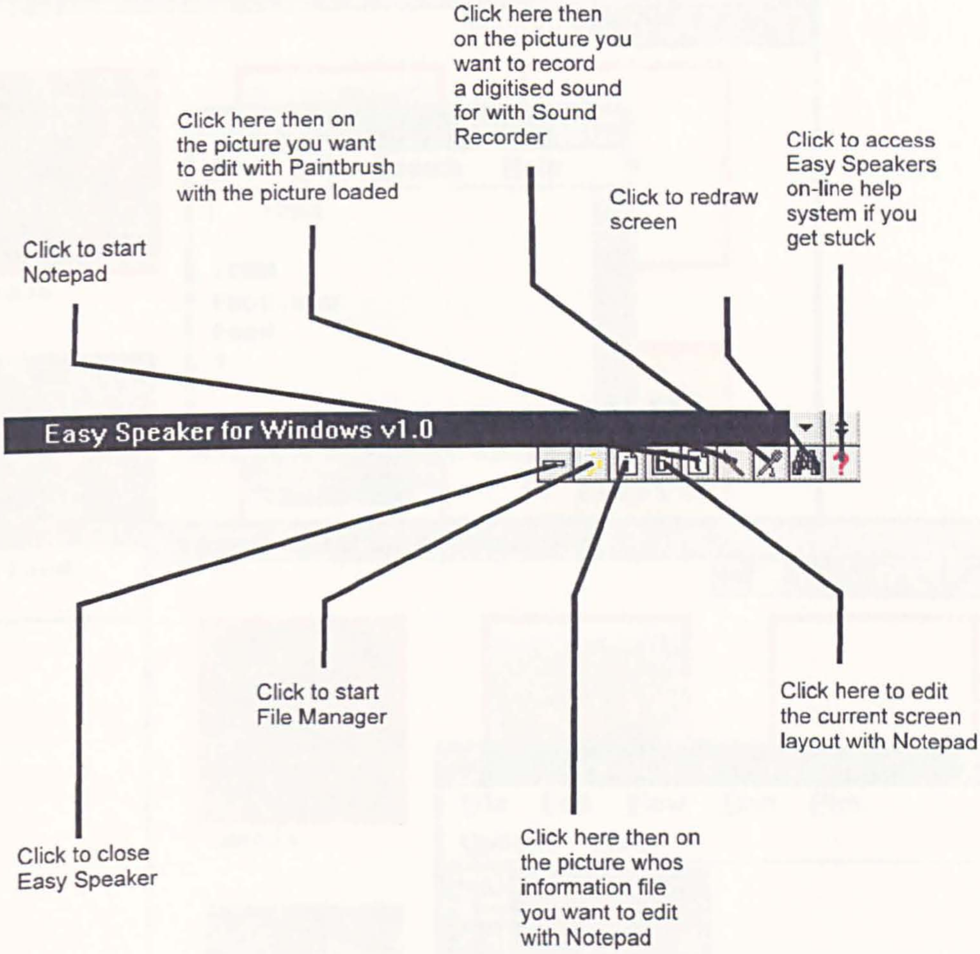
Lines 3 and 4 are crucial. If line 3 contains the uppercase word "ICON" then line 4 will be interpreted as being a digitised speech filename. When clicked on the piece of speech will be appended to the utterance under construction, and the appropriate picture will be shown at the bottom of the screen as being included in the utterance. If line 3 contains the uppercase word "LINK" then line four will be interpreted as being a screen layout filename, and the picture in question will be surrounded by a red bounding box. When clicked on, the appropriate screen layout file will be loaded and the pictures displayed. Thus the process of displaying pictures and loading information files will start a fresh.

Lines 5 and 6 describe the digitised speech which is associated with the picture. On line 5, if the picture represents speech then this is usually a straight transcription of the speech, or if a link, a description of the purpose of that link. Line 6 is simply a count of the number of words in the transcription, and hence digitised speech. If a link then the transcription length is 0. These two lines are important as they enable detailed recording of users interactions to be kept by the in-built tracking system.

Although the scripting language is very simplistic potentially huge vocabularies can be constructed that can be richly interlinked. Fortunately the size of the vocabulary and the number of links within it is unrelated to performance. In addition there is no restriction on the ratio of links to speech pictures on any screen.

In common with many object oriented programming environments Easy Speaker implemented a set of tools which allowed maintainers to treat pictures as objects in order to maintain them. By means of the maintainer toolbar, maintainers could easily edit scripts, modify pictures, and record digitised speech files. This was done by clicking on toolbar buttons and then by clicking on the symbol in question.

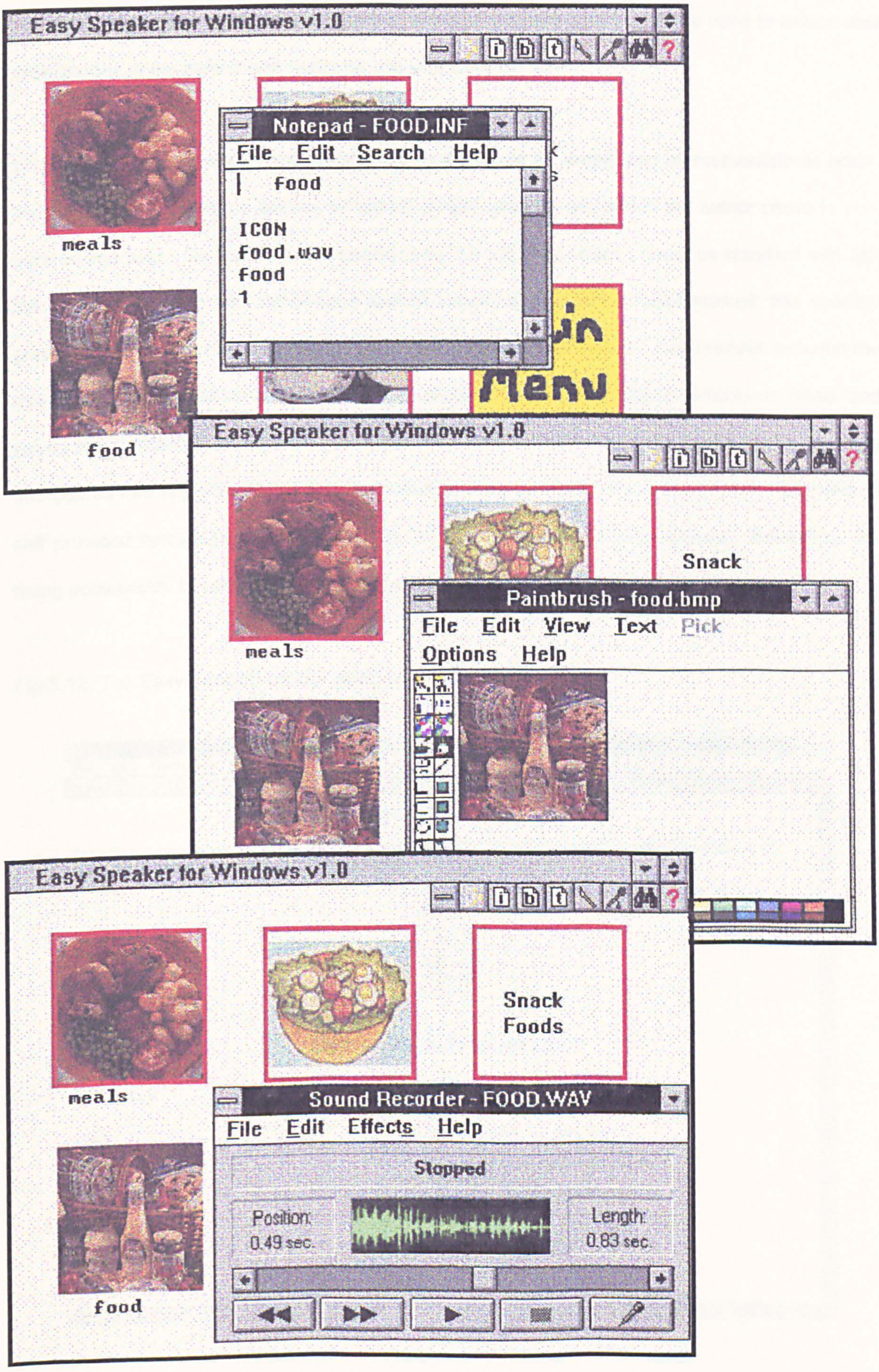
Fig 5.10, The functions of the maintainer toolbar



As Easy Speaker runs under the Microsoft Windows operating system, this enabled the author to make use of applications, or applets, that were supplied as standard with the operating system. That is, Notepad for editing ASCII text files, Paintbrush for editing pictures, Sound Recorder for recording digitised sound, and File Manager for managing files. By using familiar applets this meant that maintainers were already likely to be familiar with them, and had access to high quality documentation provided by the operating system manufacturer. Thus maintainers did not have to learn new skills and the author did not have to produce manuals for third party software tools.

The ease with which these operating system applets integrated with Easy Speaker can be seen below, where a maintainer is editing an information file for a picture with Notepad, then modifying the picture itself with Paintbrush and finally recording a piece of speech to be associated with the picture using Sound Recorder.

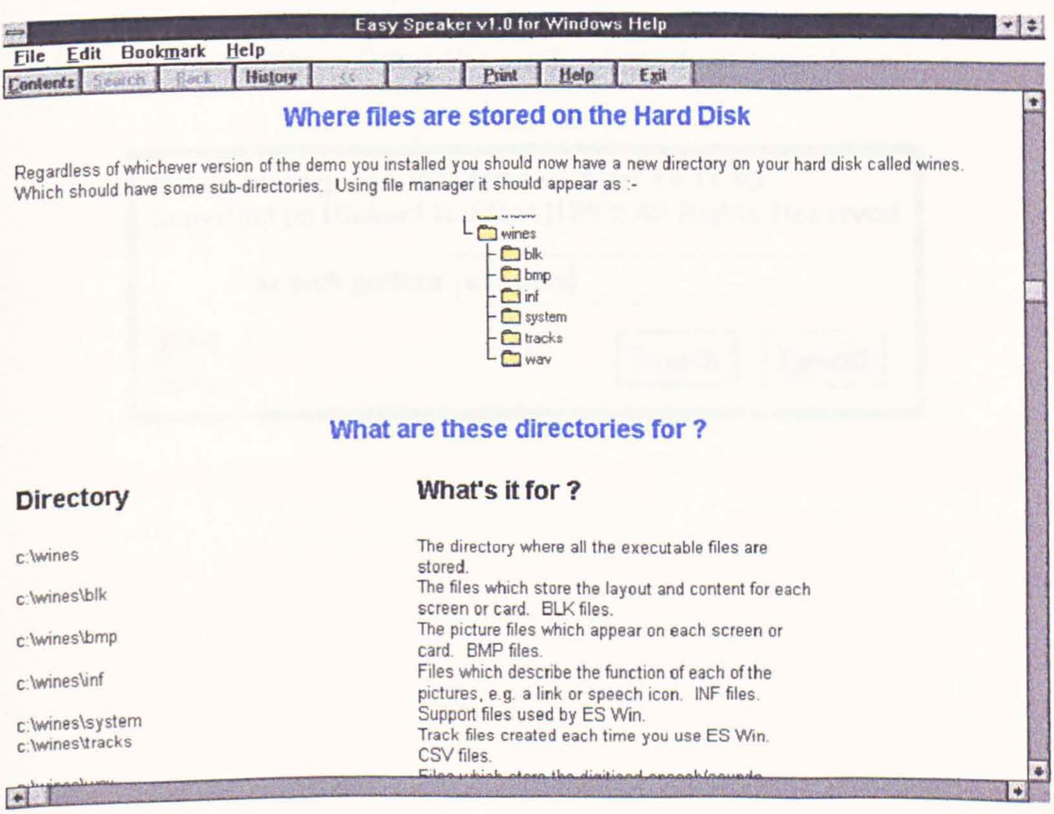
Fig 5.11, Using operating system applets to maintain Easy Speaker vocabularies



The scanning of the pictures themselves was left to maintainers who were assumed to have a MS Windows driven scanner and an appropriate piece of software which could be used to reduce and crop images to the correct size ready for use with Easy Speaker.

Despite the Easy Speaker scripting language being made as simple and as memorable as possible, and the use of system applets for editing scripts, pictures and sound the author chose to provide on-line help. On-line help was provided by the MS Help which comes as standard with MS Windows. By writing and compiling a special version of the paper based manual, this enabled graphics, searchable contents, and coloured text to be provided on-line. Key features included the use of screen grabs within the help file which showed scripts being edited, pictures modified, and sound files recorded. In addition any of the help files contents could be copied to the clipboard and pasted into any other Windows application, or they could be selectively printed. MS Help itself provided functions for hypertext jumps, pop-up definitions, printing, copying, annotating, defining bookmarks, keeping a history, and for getting help on using MS Help itself.

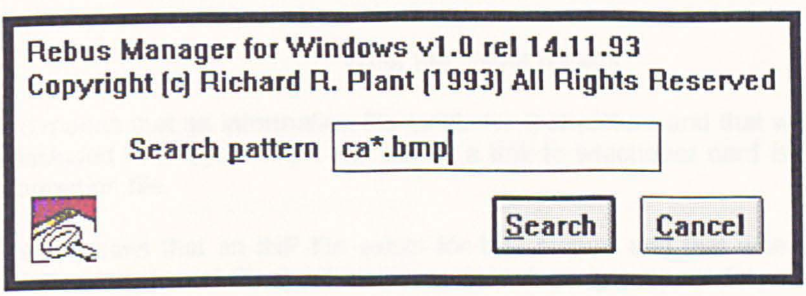
Fig 5.12, The Easy Speaker on-line guide using MS Help



Many researchers have championed the use of on-line help when combined with paper based manuals as an aid to productivity, especially for naive users, e.g. Cohill and Williges 1982, Magers 1983, and Borestein 1985. Later researchers have suggested that screen grabs, or illustrations should be used for illustrative purposes where possible, e.g. Oren et al 1990 and Baecker et al 1991. Theoretically then at least maintainers had on-line access to high quality manuals which included graphics which could be used as reference materials if they became stuck or confused.

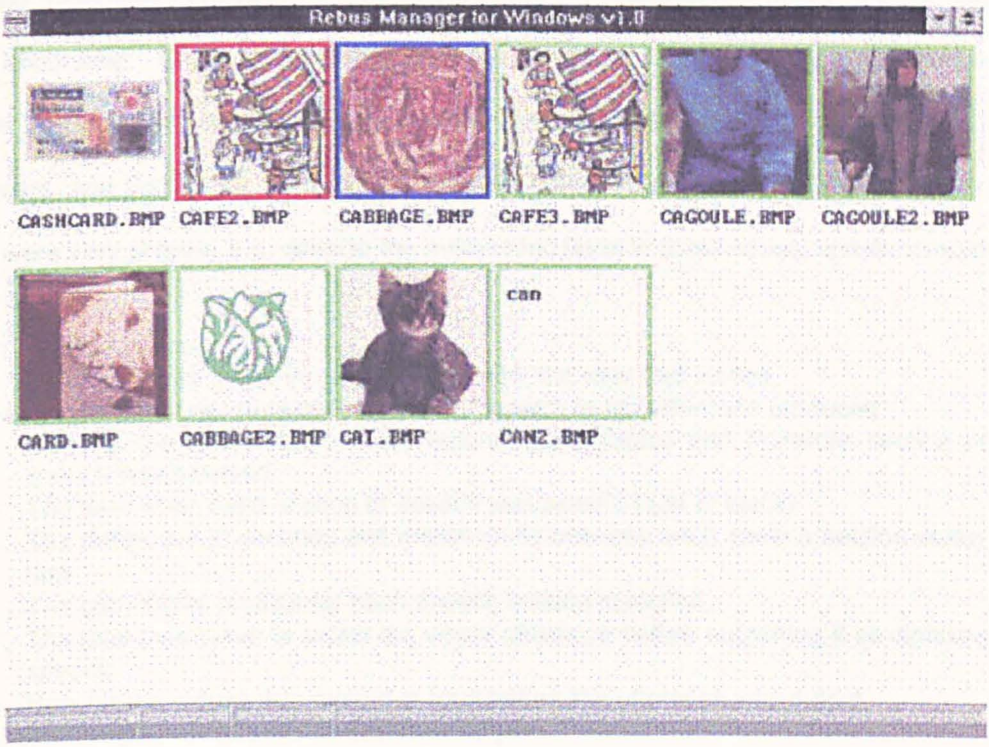
As one of the key features of Easy Speaker was that vocabularies could be as large as needed, maintainers may have difficulty in remembering what symbols they had available already. In this respect the author provided a Rebus Manager which enabled maintainers to search for pictures to be included in various part of the vocabulary. The Rebus Manager provided a way of browsing through 'albums' of pictures which are used within an Easy Speaker vocabulary. The number of pictures in the album can be narrowed down by using search strings based on the pictures file name. For example, if the maintainer wanted to know if they had a picture of a cat already on file they could enter "ca*.bmp", so that all the existing pictures beginning with the letters "ca..." would be displayed.

Fig 5.13, Using a search pattern in Rebus Manager to narrow down a search



The result of the search might look something like :-

Fig 5.14, Search results for "ca*.bmp" when looking for a picture of a cat



As an aid for the maintainer the bitmap file name is displayed under each image. This can then be typed straight into scripts. Further help as to the status of each image is provided by the bounding boxes surrounding each picture. These bounding boxes follow the colour codes :-

Colour	What the colour means
RED	Red means that an information file exists for that picture and that when that picture is included in a layout file it will act as a link to whichever card is specified in its information file.
GREEN	Green means that an INF file exists for that picture and that when that picture is included in a layout file it will act as a speech picture appending the sound specified in its information file to the utterance under construction.
BLUE	Blue means that no information file exists for that picture. You must therefore create on once you have included it within a layout file.

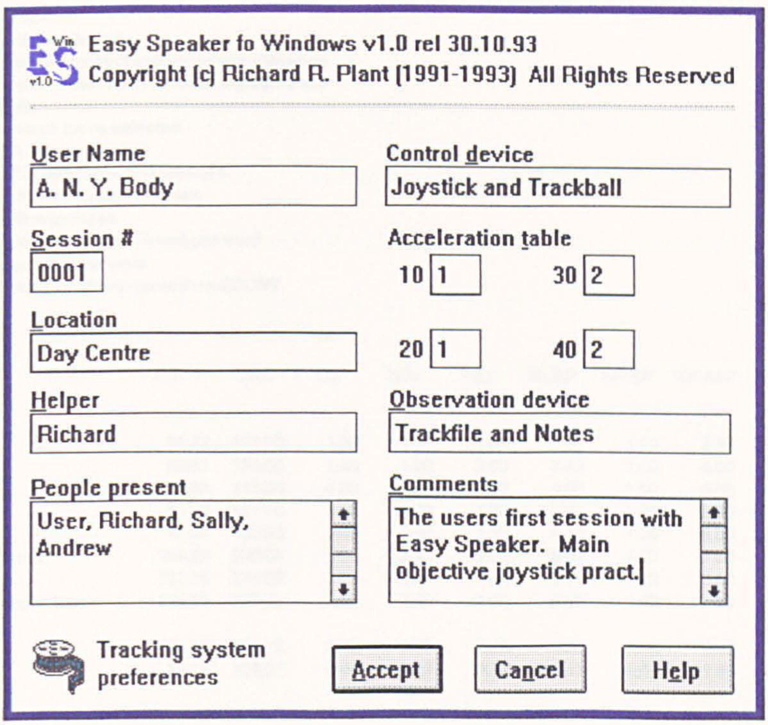
THE AUTOMATED RESPONSE TRACKING SYSTEM

Although the *good* design and implementation of Easy Speaker was considered crucial from both a users and maintainers prospective, equally important was the systems ability to track users interactions accurately. The Automated Response Tracking System (ARTS) was considered a priority and was built into the second prototype of Easy Speaker in 1990. The tracking system enabled maintainers, and the author, to measure progress objectively. Measures taken by the tracking system were very precise, e.g. down to the millisecond level in some cases, and encompassed :-

- Utterance count
- Track of where in the vocabulary hierarchy the user had visited
- The length in words of each section of speech in the utterance produced
- How far the mouse has moved before each graphic and therefore section of speech was selected
- The time when each section of speech was selected (24 hr clock)
- The difference in seconds and milliseconds between each speech section selection
- The cumulative timings for each speech section selection
- The total time taken to select the whole utterance before outputting it as digitised speech
- How far the mouse was moved in total before the utterance was output
- The length in words of the utterance output
- The number of symbols that were selected that represented speech
- The number of symbols that represented links to other parts of the vocabulary hierarchy
- Number of symbol select button presses
- Number of *say* utterance button presses
- Total number of button presses
- How far the mouse was move in order to produce each word on average
- Average number of seconds required to produce each word in the utterance
- The predicted number of words per minute based on the utterance just output

In order to automate the process as far as possible each time Easy Speaker started a dialog box prompted maintainers to fill in a few simple details. These were then used as the basis for the tracking file produced by ARTS.

Fig 5.15, The start-up dialog requesting user information for that sessions track file



Each time the system started a new track file would be started ensuring that previous versions were not overwritten or corrupted. All output from ARTS was in the form of Comma Separated Value (CSV) files which enabled a host of commercial spreadsheets and databases to read them in their native format. This ensured that maintainers could easily keep records of users progress using third party software with which they were familiar. By importing these CSV files into spread-sheet such as MS Excel this enabled progress to be plotted graphically.

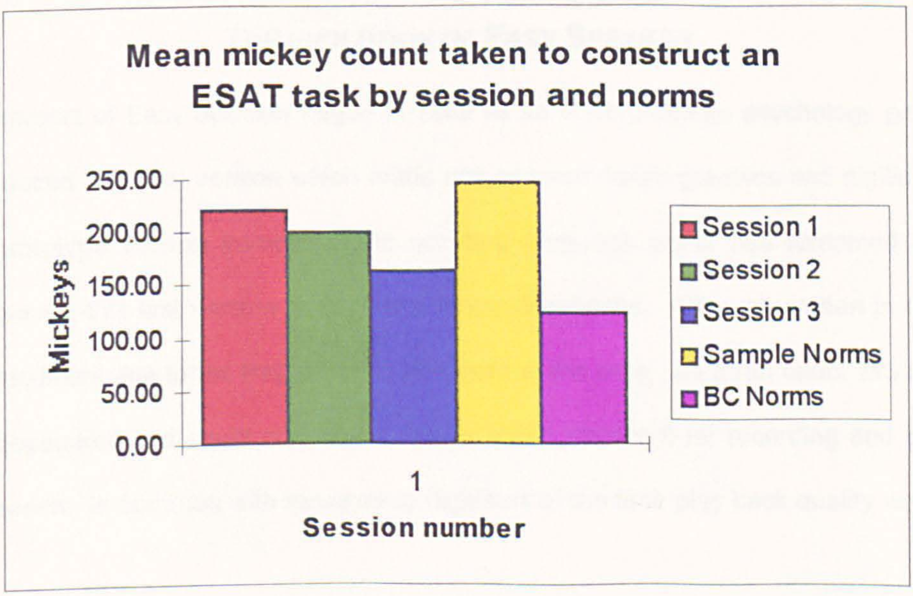
Fig 5.16, ARTS tracking data loaded into MS Excel ready for numerical analysis and plotting

ID	TFA reference ID
TRANS	Transcription of the utterence
TOT T	Total time taken to construct and repoduce utterance
TMC	Total no. of mickeys moved in constructing utterence
UL	Utterence length
NSI	Total no. of speech icons selected
NLI	Total no. of link icons selected
NLBP	Total no. of left button (selector) presses
NRBP	Total no. of right button (say) presses
TOT NBP	Total no. of button presses
M/W	Mean no. of mickeys (approx 1mm) per word
SEC/W	Mean no. of seconds per word
W/MIN	Predicted Words Per Minute based on SEC/W

Actual Results

	TRANS	TOT T	TMC	UL	NSI	NLI	NLBP	NRBP	TOT NBP	M/W	SEC/W	W/MIN
Assessment session 1												
1.00	meals	60.42	192.00	1.00	1.00	1.00	2.00	1.00	3.00	192.00	60.42	0.99
2.00	summer	62.61	123.00	1.00	1.00	2.00	3.00	1.00	4.00	123.00	62.61	0.96
3.00	trainers	65.52	113.00	1.00	1.00	3.00	4.00	1.00	5.00	113.00	65.52	0.92
4.00	muscles	89.25	157.00	1.00	1.00	4.00	5.00	1.00	6.00	157.00	89.25	0.67
5.00	ham/salad	98.48	450.00	2.00	2.00	2.00	4.00	1.00	5.00	225.00	49.24	1.22
6.00	read/library/book	106.89	200.00	3.00	3.00	3.00	6.00	1.00	7.00	50.00	26.72	1.68
7.00	is it/raining	132.26	274.00	3.00	2.00	3.00	5.00	1.00	6.00	91.33	44.09	1.36
8.00	get the nurse/sore/back	124.96	257.00	5.00	3.00	3.00	6.00	1.00	7.00	51.40	24.99	2.40
Ave		92.55	220.75	2.13	1.75	2.63	4.38	1.00	5.38	125.34	52.86	1.28
SD		28.13	109.02	1.46	0.89	0.92	1.41	0.00	1.41	63.06	21.34	0.55

Fig 5.17, Graphical representation of ARTS data can easily be accomplished within most spreadsheets



The ability to import the track files into other standard packages such as, spreadsheets, databases and word processors was considered crucial. For example, maintainers could potentially drop a track file, and accompanying graph, straight into a word processed report of a users progress. The use of bespoke evaluation materials including the potential for track files is discussed more fully in chapter seven.

In sum the Easy Speaker scripting language was made as simple as possible, and an object oriented approach was followed to allow for easy editing of scripts, images and sounds by means of standard operating system applets. If maintainers became stuck or confused they could refer to the high quality on-line guide, and if necessary print sections of it out "on-the-hoof". Other tools such as the Rebus Manager were provided to allow collections of digitised pictures to be easily maintained and utilised. The Automated Response Tracking System (ARTS) was considered crucial as this enabled users interactions with the system to be recorded verbatim and objectively assessed at a later date. Thus progress could be measured over time.

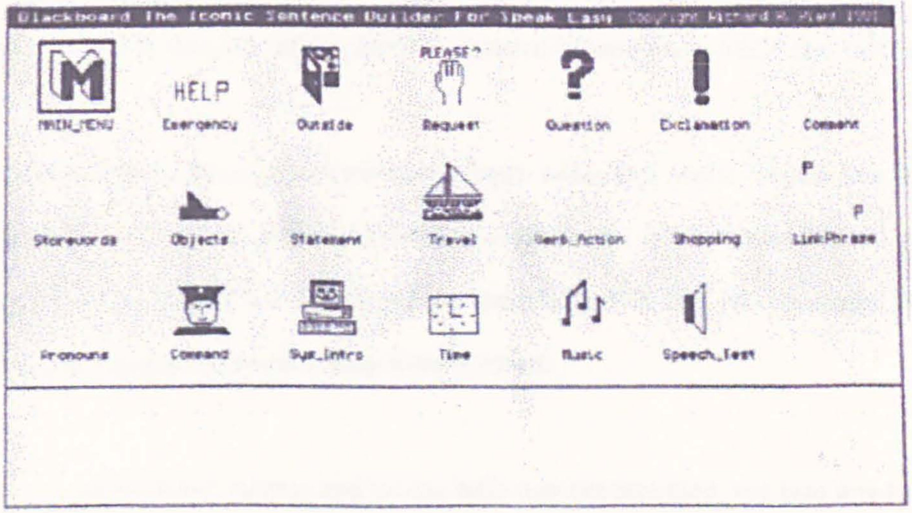
Having outlined the design ethos, and functionality of the current version of Easy Speaker from both a user and maintainers prospective it is useful to briefly consider the earlier prototypes that preceded it.

THE EVOLUTION OF EASY SPEAKER

The development of Easy Speaker began in 1989 as an undergraduate psychology project. This project produced an initial version which made use of hand drawn graphics and digitised speech. This first prototype version introduced the scripting language which has remained unchanged since. However, this first version had several major drawbacks. Although written in C, performance was mediocre due to the limitations of hardware at the time. As it ran under MS DOS it was hardware dependent and could only use a Covox Voice Master II for recording and playback of digitised speech. In common with many other digitisers of the time play back quality was relatively poor.

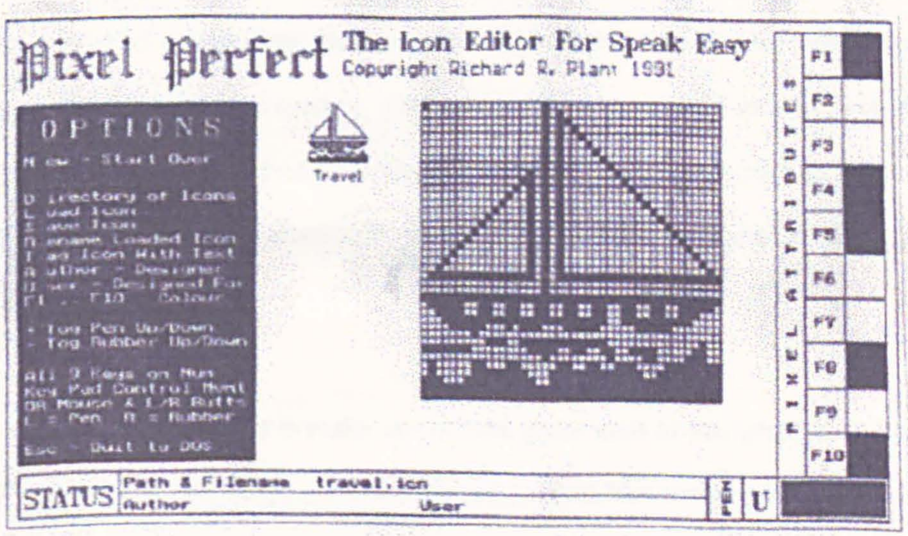
The actual user interface was very similar to the current version of Easy Speaker with red bounding boxes indicating links, and unbounded graphics speech. The following screen grab shows the commonality of the interface.

Fig 5.18, The first prototype version of Easy Speaker



Due to the limitations imposed by DOS, graphics could only be shown in 16 fixed colours. All graphics were hand drawn and were in a non-industry standard graphics format developed by the author. As a result graphics could only be edited using another program (Pixel Perfect) developed by the author. The size of each graphic was fixed 96x96 pixels.

Fig 5.19, Editing symbols using Pixel Perfect



The screen resolution throughout was standard VGA, 640x480 pixels in 16 colours.

In this first version there was no tracking (ARTS) system, and so any observations were left to the maintainer. In addition all maintenance of scripts, screen layout, and information files was left to the maintainer. That is, they had to use their own choice of third party ASCII text editing tools.

The user controlled Easy Speaker via the cursor keys, enter and space. Again due to the limitations imposed by DOS a changeable mouse-type cursor could not be implemented. As a result the currently selected symbol was signified by a black bounding box which moved incrementally from symbol to symbol as movement keys were pressed.

There was no error message facility, and on-line help was not provided; nor was any form of paper based manual.

The second Easy Speaker prototype was again DOS based and made use of the Covox digitising hardware. However, this time several supplementary programs were developed which would allow conversion between industry standard graphic and sound formats. This meant that BMP graphics files and WAV sound files were supported. In addition any other Pulse Code Modulation (PCM) sound files were also supported.

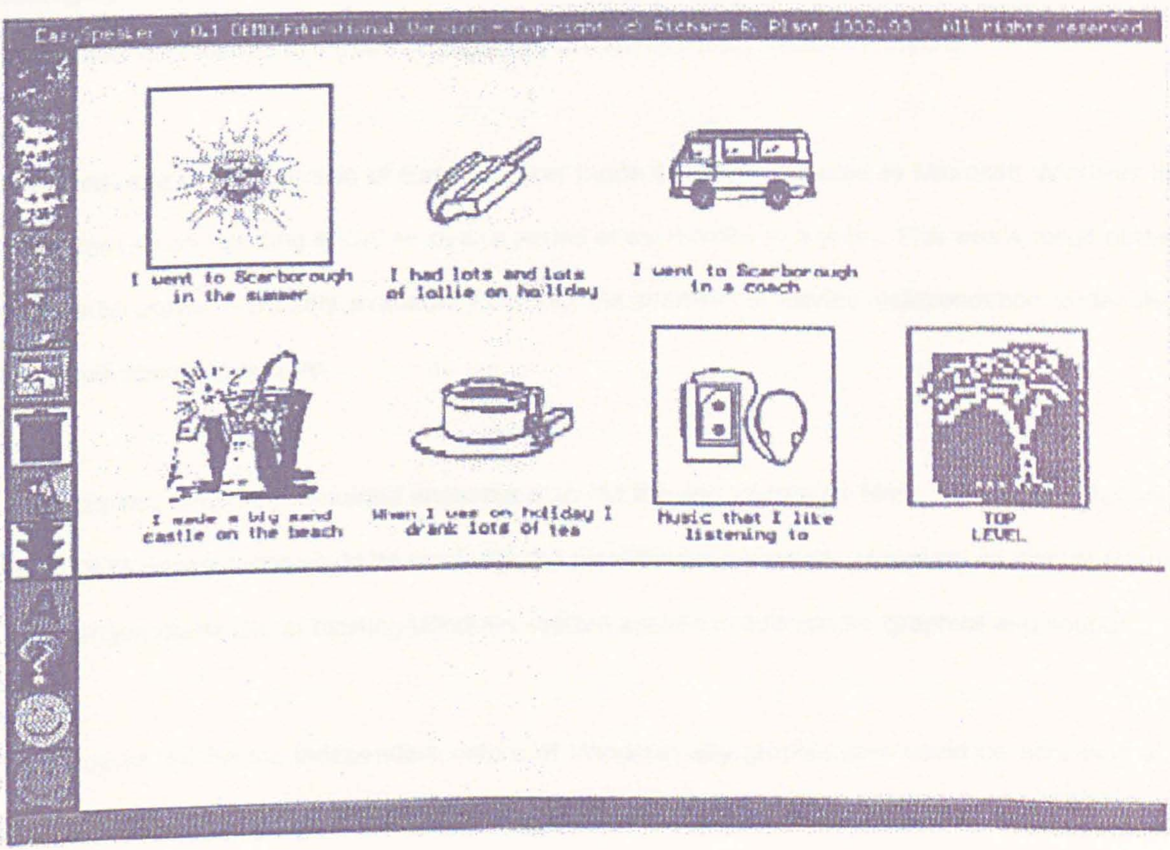
Although the quality of sound files could now be up to 8 bit, 22,000 Hz, mono, the quality of graphics was still restricted to 16 colours. Despite relatively low quality graphic reproduction as a result of the ability to display industry standard graphic files, scanned images of people, objects and places, could now be used. Albeit if they were dithered down to 16 colours taken from a fixed colour palette.

Again maintainers were expected to make use of third party tools to edit graphic, sound and ASCII text script files. The scripting language and other associated files remained unchanged. The only tools that were provided were those for converting between various formats of graphics and sound files.

The second prototype introduced the concept of the tracking system, or ARTS, as a way of recording a users interactions for analysis at a later date. The content of the tracking system was very similar to that implemented by the current version of Easy Speaker. This file was in stored as an industry standard Comma Separated Value (CSV) file and could be imported into most standard spreadsheets for analysis.

From a users point-of-view the interface had changed substantially with the addition of a fast speech toolbar. This enabled smaller 32x32, 16 colour icons to be displayed. These were in the MS Windows ICO format enabling the use of existing icon libraries. The fast speech toolbar in common with the current version of Easy Speaker remained constant regardless of whatever screen the user was on, and allowed access to common words or phrases.

Fig 5.20, The second prototype version of Easy Speaker



The representation of the cursor remained constant. However in addition to cursor control, mouse movement was also supported. This meant that any serial device could be supported using the standard PC serial ports. Due to the ranges of movement shown by some users the concept of acceleration tables were also implemented. This meant that a maintainer could set the mouse movement ratio to on screen cursor movement. Thus the effects of tremor and poor motor control could be accounted for when using analogue control devices. In addition a digital port was provided by directly interfacing to the keyboard. This enabled digital joystick and other devices to control Easy Speaker.

To accompany this version a 32 page maintainer manual was produced, and rudimentary error messages were given when script errors etc... were found. On-line help was still not provided.

This was the first version that was trialed with real users. Results with real users were very encouraging and prompted further development of the software and more thorough user trials. For a discussion see Rostron and Plant (1992) and Rostron, Plant and Hermann (1994).

The third, and current version of Easy Speaker made the transformation to Microsoft Windows in 1993 after much rewriting in C/C++ over a period of six months to a year. This was a result of the increasing power of readily available PCs and the promise of device independence under the Windows operating system.

The scripting language remained unchanged as did the use of related files. This meant that existing DOS vocabularies could be used with the new Windows version. However, as discussed the new version made use of existing Windows system applets to edit scripts, graphics and sounds.

By virtue of the device independent nature of Windows any graphic card could be accessed via high level operating system calls. This meant that for the first time Easy Speaker could make use of photorealistic images in up to 16.7 million colours. In addition sound up to CD quality could be supported by using high level system calls. In short any devices that ran under Windows could be used with Easy Speaker. This meant that any brand of graphics card capable of true colour could be used, and any sound card capable of recording and playback under Windows was adequate.

In terms of input device again any serial device could be used, with the usual acceleration table to compensate for the effects of poor motor control. The Windows environment had the added attraction that it enabled any device that replaced the mouse could be used to control Easy Speaker. This meant that touch screens etc... could be utilised without modification to the Easy Speaker code as they simply replaced the mouse under Windows.

The actual pointer the user saw was now free ranging and could be selected as desired by using a third party cursor changing program. This meant that the movement of the cursor now took an analogue form, rather than being a black bounding box jumping from symbol to symbol when moved.

Within this version the tracking system was expanded to encompass every aspect of user interaction with the system. Again this was recorded in the standard CSV format ready to be loaded in to any spreadsheet running under Windows.

A sixty page maintainer manual was produced that enabled maintainers to produce their first vocabularies with relative ease, and act as a reference for the more experienced. To supplement this full on-line help was provided by making use of the Windows Help system. Both the paper based version and the on-line guide contained full colour graphics.

In order to help maintainers in constructing and maintaining vocabularies full error messages were provided. These highlighted any errors in scripts or related files as Easy Speaker ran, enabling them to be fixed interactively. Combined with the object-orientated design this made the current version of Easy Speaker considerably easier to use than previous versions.

In addition various utilities were written by the author to catalogue and search through the potentially huge number of digitised pictures a maintainer may have built up. One such utility was the

Rebus Manager discussed earlier. Thus combined with Easy Speaker itself and operating system applets maintainers could easily and quickly construct high quality hierarchies for specific users.

SUMMARY

Specific design guidelines followed during the development of Easy Speaker have been discussed in relation to potential users. These centre around the application of the SSOA model in relation to direct manipulation, and the relevance of semiotics in symbol layout and design.

Just as importantly the design guidelines in relation to maintainers, and the maintainer and scripting interface of Easy Speaker have been discussed. This focused on how actual hierarchies were constructed by means of utilising operating system applets and how these eventually appeared on the screen.

The operation and analysis of ARTS tracking data produced by actual users was discussed in relation to overall design objectives. The evolution of Easy Speaker from its inception was discussed. Improvements between versions were noted as was the types of hardware required to make full use of the system.

However, despite theoretically good design, and improvements over previous versions it was not demonstrably known whether real maintainers could maintain an Easy Speaker hierarchy satisfactorily. This question is addressed in the next chapter in which three real maintainers are questioned on their opinions toward Easy Speaker using the University of Maryland Questionnaire for User Interface Satisfaction (QUIS). The question of whether real users could make productive use of the system by virtue of good design is formally assessed in chapter nine. In chapter ten off-device improvements in language and cognition are considered.

CHAPTER 6: Can maintainers maintain Easy Speaker hierarchies ?

Although the requirements of potential users are paramount so too are the needs of potential facilitators, caregivers, or health care professionals who might set-up and maintain an Easy Speaker hierarchy for a given user, or users.

Despite Easy Speaker being designed to be simple for both users and maintainers alike, the opinions and attitudes of real maintainers to it was unknown. In order to assess the opinions and attitudes toward the maintenance of Easy Speaker hierarchies the University of Maryland Questionnaire for User Interface Satisfaction 5.0 Long (QUIS 5.0L) was employed.

This was administered to novice maintainers, and then again after a week of utilising Easy Speaker. This was done to see if initial impressions lasted, or whether these became more or less favourable. Results could then be used to consolidate liked features and to improve disliked ones.

Also as the QUIS is based on a Likert scale, with a known mean response, maintainers responses were compared with the expected mean. This was done to evaluate whether maintainers thought Easy Speaker was better, or worse, than the expected average.

Suggestions as to what might be changed to improve maintainers views, as assessed by the QUIS, are also discussed.

INTRODUCTION

With a piece of software such as Easy Speaker the maintenance of the hierarchies, or vocabularies, that it is built around is crucial to its success. As the user develops so too should the complexity and depth of the vocabulary provided. Although on prima facie evidence the simplicity of the scripting language, object oriented design and use of common third party applets should not provide a barrier to maintenance, this has not been collaborated by actual maintainers.

In time pressured service environments such as the NHS, the ease with which hierarchies can be maintained is crucial. Regardless of initial enthusiasm, if potential maintainers find maintenance chores too demanding, or time consuming, they are unlikely to be carried out. Unfortunately this is usually to the detriment of the user. Initial attitudes toward Easy Speaker are crucial.

Whether potential maintainers think it is better, or worse, than average as a piece of software is essential. If initial opinions and attitudes are negative then these are likely to colour future use and maintenance of the system. In addition to assessing initial opinions it is also crucial to monitor how opinions change over a short period of time in which the maintainer changes from being novice to becoming what might be considered expert. On becoming an expert do they consider the system to be better or worse than when they started out ? If they consider the system to be worse then again this is likely to colour their use of it in a negative way. If on the other hand they consider the system to be better then this can only have a positive effect. Both initial and later more experienced views are important for determining whether Easy Speaker can be maintained by maintainers. Strengths and weaknesses are also highlighted so they might be paid attention to in future versions.

CAN MAINTAINERS MAINTAIN THE SYSTEM ?: AN EXPERIMENTAL EVALUATION

The paradigm

In order to assess maintainers attitudes and experience with Easy Speaker three real maintainers were asked to complete a University of Maryland Questionnaire for User Interface Satisfaction 5.0L, after just a few hours of experience (novice) and then again after a week of building Easy Speaker vocabularies (experienced). This was done to see if initial opinions lasted, or whether they became more or less favourable toward the maintenance of Easy Speaker vocabularies. These opinions would be used to assess how well the maintainer interface and scripting language were designed, and as a result be used in the design of future versions. In addition ratings would be compared with the expected rating for each QUIS item. This would be used to assess whether maintainers ratings were better, or worse, than the expected average.

Design

A fully systematic repeated measures design was employed as each maintainer would be administered a QUIS when novice and then again when considered an expert.

Subjects

All three maintainers were using Easy Speaker with real users as part of their final year undergraduate psychology project. Two were female, and one was male. Collectively they shared a mean age of 27 years, and all had some previous computer experience.

To give some idea of maintainers previous computer experience the following table can be constructed from initial (novice) QUIS questionnaires :-

Fig 6.1, Previous computer experience of maintainers

Personal Computer	All
Word Processing	All
Games	All
Keyboard	All
Mouse	All
Colour monitor	All
Floppy disks	All
Hard Disks	All
Compact Disk drives	All

All maintainers had experience in using word processors on Personal Computers. All had previous experience in playing games on computers. All could, as expected, use a keyboard and mouse as means of input. All had used computers equipped with colour monitors, floppy and hard disks and CD ROMs. Although these maintainers were University students one might expect this pattern of computer experience to be common in those groups likely to act as maintainers for Easy Speaker vocabularies in service environments. Namely NHS staff, residential caregivers and enthusiastic parents. They might be termed, for sake of argument, novice computer users. It must be remembered however that University students are not mirror images of caregivers in the real world.

Apparatus

Six copies of the University of Maryland's Questionnaire for User Interface Satisfaction version five, long form developed by Shneiderman and Norman (1988-1991). Based upon a simple 'form fill-in' Likert scale the QUIS questioned maintainers on aspects of seven areas :-

- ❶ Type of system to be rated
- ❷ Past experience
- ❸ Overall user reactions
- ❹ Screen
- ❺ Terminology and system information
- ❻ Learning
- ❼ System capabilities

As an aid in rating each item had a semantic differential scale which accompanied the standard numerical Likert scale. Fortunately throughout the QUIS all the positive attributes are biased toward the right of the scale as the example below shows. Raters were simply required to circle the desired value on the scale, or circle "NA" if they thought the question was Not Applicable, or they couldn't answer.

Fig 6.2, An example QUIS item from the "Screen" section

	Illogical							Logical		
Arrangement of information on the screen	1	2	3	4	5	6	7	8	9	NA

As the scale ranges from 1 to 9, one can calculate the expected response to be 5, which represents an average opinion for that item.

Standard IBM PC compatible computer equipment and software was used throughout. All maintainers made use of the following pieces of hardware and software :-

- Elonex 486 sx 25 Mhz with 8 Mb of RAM connected to a 17" Super VGA monitor
- Running Microsoft Windows version 3.1
- A copy of Easy Speaker for Windows version 1.0
- A copy of Microsoft Excel spreadsheet for analysing ARTS track files
- A LogiTech Audioman attached to the parallel port for sound recording and reproduction
- A LogiTech Scanman hand held true colour scanner for scanning images
- A Hewlett Packard Deskjet 550c colour printer for printing screen layouts etc...
- A standard mouse and keyboard was used for all input

Procedure

Maintainers were individually instructed in the operation of Easy Speaker and the scripting language it employs over the period of approximately one hour. During this initial period they were allowed to interact with the system and to question the author as and when they desired. At the end of this initial one hour session they were given the 60 page instruction manual and were left to practice what they had learnt. After a further hour they were asked to complete the QUIS. At this stage the maintainer was considered as being novice.

Over the period of approximately a week each maintainer had the opportunity to construct hierarchies for the user, or users, they would be working with. This involved scanning images, editing graphics, writing scripts, recording sound files etc... At any time they were encouraged to ask the author questions, consult the paper based manual, or make use of the on-line help system.
During this week all maintainers "clocked-up" well over 10 hours experience with Easy Speaker. At the end of this initial week a second copy of the QUIS was administered. At this stage the maintainer was considered as being an *expert*.

All three maintainers underwent the process described in isolation during a week long timeslot within a three month period. After this initial week maintainers continued to use Easy Speaker and to construct hierarchies for their intended real users who would be subjects in each of their final year undergraduate projects.

RESULTS

In terms of software design, information from four sections of the QUIS are crucial. Namely the sections dealing with the screen, terminology and system information, learning, and system capabilities. As all these sections are based upon a nine point Likert scale this enables statistical comparison of the actual responses of maintainers with those which one might expect. That is, as the scale ranges from 1, least favourable, to 9, very favourable, it is reasonable to suppose a system of average design would receive a mean of 5 for each section. A higher mean would suggest a better than average system, any lower, a poorly designed system. The deviation from the expected value can be given a statistical significance by applying a single sample t-test. This can be done for both novice and experienced maintainers.

Changes in opinion brought about by experience with the system can also be assessed by applying a dependent t-test to assess the difference between novice and experienced maintainers ratings. For example, an initially favourable rating for a feature, may be rated as being poorer as the maintainer becomes more experienced and may discover a short-coming of the particular feature. For the system to be considered as success, ratings should rise significantly as maintainers become more experienced. At worst they should remain constant.

The screen

The screen section of the QUIS contained questions that asked maintainers to rate aspects of the display such as the "Amount of information that can be displayed on a screen". A complete list of questions from the screen section are shown below with their accompanying mean and the verbal description of the Likert scale.

Fig 6.3, Maintainers QUIS responses for the screen design aspects of Easy Speaker

Item	Likert scale value 1 - 9	Novice Mean	Novice SD	Expert Mean	Expert SD
Characters on the computer screen	hard to read - easy to read	8.33	1.15	8.00	1.00
Image of characters	fuzzy - sharp	5.67	4.16	5.67	4.16
Character shapes (fonts)	barely legible - very legible	7.33	1.53	8.33	1.15
Was the highlighting on the screen helpful	not at all - very much	5.00	-	-	-
Use of reverse video	unhelpful - helpful	-	-	-	-
Use of blinking	unhelpful - helpful	-	-	-	-
Were the screen layouts helpful	never - always	5.00	0.00	4.50	0.71
Amount of information that can be displayed on screen	inadequate - adequate	3.50	3.54	3.50	3.54
Arrangement of information on screen	illogical - logical	7.00	0.00	6.50	2.12
Sequence of screens	confusing - clear	7.00	1.73	7.67	1.53
Next in a sequence	unpredictable - predictable	7.67	1.53	8.00	1.73
Going back to the previous screen	impossible - easy	4.33	3.06	5.33	3.79
Beginning, middle and end of tasks	confusing - clearly marked	-	-	-	-
	Mean	6.33		6.58	
	SD	2.46		2.60	

* N.B. Overall means are calculated from the raw data

\- Could not be computed

Novice maintainers found Easy Speaker screen attributes to be significantly better than average (t=2.64, df=24, sig=0.01), when compared with an overall expected mean of 5. When they became classed as expert maintainers this position was strengthened, (t=2.98, df=23, sig=0.005).

Although Easy Speakers screen attributes were rated as being better than average by both novice and expert maintainers, there was no significant change in ratings between users being novice and experienced. This suggests that their attitude toward the screen layout remained constant regardless of level experience. It became neither more or less favourable.

Terminology and system information

Fig 6.4, Maintainers QUIS responses for the terminology and system information aspects of Easy Speaker

Item	Likert scale value 1 - 9	Novice Mean	Novice SD	Expert Mean	Expert SD
Use of terms throughout system	inconsistent - consistent	6.50	3.54	7.50	2.12
Task terms	inconsistent - consistent	4.50	0.71	7.50	2.12
Computer terms	inconsistent - consistent	4.50	0.71	7.50	2.12
Does the terminology relate well to the work you are doing	unrelated - related	3.50	0.71	5.50	2.12
Computer technology is used	too frequently - appropriately	4.50	0.71	5.00	1.41
Terms on screen	ambiguous - precise	3.50	2.12	6.50	2.12
Messages which appear on screen	inconsistent - consistent	4.00	0.00	5.50	0.71
Position of instructions on screen	inconsistent - consistent	7.00	2.83	7.00	2.83
Messages which appear on screen	confusing - clear	3.00	1.41	6.00	1.41
Instructions for commands and choices	confusing - clear	3.00	1.41	6.50	3.54
Instructions for correcting errors	confusing - clear	3.00	0.00	4.00	0.00
Does the computer keep you informed about what it is doing	never - always	4.00	0.00	4.00	0.00
Performing an operation leads to a predictable result	never - always	3.50	0.71	6.50	0.71
User can control amount of feedback	never - always	2.50	0.71	4.00	1.41
Error messages	unhelpful - helpful	3.00	-	4.00	1.41
Error messages clarify the problem	never - always	2.00	1.41	4.00	1.41
Phrasing of error messages	unpleasant - pleasant	3.00	2.83	4.00	1.41
	Mean	3.85		5.67	
	SD	1.73		1.87	

* N.B. Overall means are calculated from the raw data

- Could not be computed

In terms of terminology and system information the QUIS highlighted many areas for concern for novice maintainers. Almost all of the ratings given by novice maintainers were significantly below average (t=3.81, df=32, sig=0.0005). Furthermore, even after becoming experienced there was no significant improvement over the expected average rating (t=1.53, df=32, NS). This suggests that maintainers, despite the simplicity of the scripting language and maintainer interface initially found Easy Speaker relatively difficult to maintain. When experienced the situation improved somewhat, but still ratings were only average.

Despite initially below average ratings, maintainers ratings did significantly improve as they became more expert in constructing Easy Speaker vocabularies (t=5.66, df=32, sig 0.0005). However, expert ratings only rose to around average.

Learning

Fig 6.5, Maintainers QUIS responses for learning to use Easy Speaker

Item	Likert scale value 1 - 9	Novice Mean	Novice SD	Expert Mean	Expert SD
Learning to operate the system	difficult -easy	3.33	1.15	6.50	3.54
Getting started	difficult -easy	3.33	1.15	6.50	3.54
Learning advanced features	difficult -easy	3.50	0.71	9.00	-
Time to learn to use the system	slow - fast	6.00	2.65	7.00	2.83
Exploration of features by trial and error	discouraging - encouraging	4.67	0.58	6.00	-
Exploration of features	risky - safe	4.33	2.08	3.50	3.54
Discovering new features	difficult - easy	6.00	1.41	4.00	4.24
Remembering names and use of commands	difficult - easy	5.67	1.53	8.00	1.00
Remembering specific rules about entering commands	difficult - easy	5.33	2.08	8.33	1.15
Can tasks be performed in a straight forward manner	never - always	6.67	2.08	7.67	1.53
Number of steps per task	too many - just right	5.67	1.53	5.00	2.65
Steps to complete a task follow a logical sequence	rarely - always	6.00	1.73	7.00	2.65
Completion of sequence of steps	unclear - clear	6.00	2.65	7.00	2.65
Help messages on screen	confusing - clear	4.50	2.12	4.00	-
Accessing help messages	difficult - easy	3.50	0.71	7.00	-
Content of help messages	confusing - clear	4.00	1.41	5.00	-
Amount of help	inadequate - adequate	3.50	0.71	4.00	-
Supplemental reference materials	confusing - clear	5.00	3.61	6.00	2.65
Tutorials for beginners	confusing - clear	2.00	-	6.00	-
Reference manuals	confusing - clear	6.50	3.54	6.50	3.54
	Mean	5.33		6.40	
	SD	1.95		2.39	

* N.B. Overall means are calculated from the raw data

\- Could not be computed

As suggested by ratings on terminology and system information novice maintainers rated some aspects of Easy Speaker as being quite difficult to learn. Low ratings focused around starting to use the system, the amount of help provided and tutorials for beginners.

Despite this apprehension novice maintainers rated Easy Speaker of being of average difficulty to learn to use productively (t=0.28, df=50, NS). As they became more expert they rated Easy Speaker as easier then average to learn (t=3.70, df=39, sig=0.0005).

Differences between novice and expert maintainers ratings were, as expected, significant (t=2.88, df=39, sig=0.005). The more expert maintainers became, the more highly they rated the system.

System capabilities

Fig 6.7, Maintainers QUIS responses for the system capabilities of Easy Speaker

Item	Likert scale value 1 - 9	Novice Mean	Novice SD	Expert Mean	Expert SD
System speed	too slow - fast enough	5.67	2.31	6.33	1.15
Response time for most operations	too slow - fast enough	5.67	2.31	6.00	1.00
Rate information displayed	too slow - fast enough	5.33	1.53	5.67	1.15
How reliable is the system	unreliable - reliable	4.00	1.00	6.33	1.15
Operations are	undependable - dependable	5.00	1.00	6.33	0.58
System failures occur	frequently - seldom	4.33	0.58	6.33	2.08
System warns the user about potential problems	never - always	5.50	2.12	4.50	0.71
Systems tends to be	noisy - quiet	7.67	2.31	7.67	2.31
Mechanical devices such as fans, disks and printers	noisy - quiet	8.00	1.73	7.67	2.31
Computer tones, beeps, clicks, etc...	annoying - pleasant	6.00	1.00	8.33	1.15
Correcting your mistakes	difficult - easy	4.67	0.58	7.67	1.53
Correcting typos or mistakes	complex - simple	5.33	1.53	7.33	2.08
Ability to undo operations	inadequate - adequate	4.67	3.21	7.33	2.08
Are the needs of both experienced and inexperienced users taken into consideration	never - always	5.00	0.00	5.67	1.15
Novices can accomplish tasks knowing only a few commands	with difficulty - easily	3.67	2.89	5.33	3.21
Experts can use features/shortcuts	with difficulty - easily	6.50	3.54	5.33	2.52
	Mean	5.42		6.56	
	SD	1.97		1.83	

* N.B. Overall means are calculated from the raw data

Novice maintainers rated some aspects of Easy Speaker system capability as being below average. Three critical ratings concerned system reliability, and the ability for novices to accomplish task knowing only a few commands.

Concerns such as these meant that maintainers rated system capabilities as being only average (t=1.44, df=44, NS). However, as they became more expert they rated system capability as being significantly above average (t=5.82, df=46, sig=0.0005).

Differences between novice and expert maintainers ratings were, as expected, significant (t=3.97, df=44, sig=0.0005) suggesting the more they used the system the more capable they found it to be.

Overall user reactions to the system

In addition to the more objective sections of the QUIS there was an opportunity for maintainers to subjectively rate their impressions of using the system. Again this was based on a Likert scale although this time there was no question to answer, but simply a semantic scale on which to register their subjective impression.

Fig 6.8, Maintainers QUIS responses for overall user reactions to Easy Speaker

Likert scale value 1 - 9	Novice Mean	Novice SD	Expert Mean	Expert SD
terrible - wonderful	4.67	1.53	6.67	0.58
frustrating - satisfying	3.33	1.53	7.00	0.00
dull-stimulating	6.33	1.53	7.00	0.00
difficult - easy	3.33	0.58	7.67	1.53
Inadequate power - adequate power	5.67	0.58	5.33	1.15
rigid - flexible	5.33	3.06	6.00	2.00
Mean	4.78		6.61	
SD	1.83		1.24	

* N.B. Overall means are calculated from the raw data

Initial ratings suggested that novice maintainers overall attitude towards Easy Speaker was about average (t=0.51, df=17, NS). However, novice users found it difficult and frustrating to use, and rated it as so on those two items. When expert however ratings were significantly above average (t=5.50, df=17, sig=0.0005).

Differences between novice and expert maintainers ratings were again significant (t=3.72, df=17, sig=0.001) suggesting the more they used the system the more they liked it. Interestingly the two areas in which overall reactions were rated lowest by novice users were rated highest when expert. That is, now instead of finding Easy Speaker difficult and frustrating to use they found it easier and more satisfying.

DISCUSSION

Although Easy Speaker was in its first version in the Windows environment initial attitudes and opinions of real maintainers was encouraging. Real in terms of the fact that they were involved in projects that required them to produce functional hierarchies for actual users who suffered from many of the linguistic and cognitive deficits discussed in chapter two.

Although maintainers were what might be considered novice computer users all managed to produce usable hierarchies for their intended user(s), and went on to use those hierarchies with them. One particularly successful application of one of the three maintainers was with an aphasic non-verbal adult. For a full discussion see Rostron, Ward and Plant (1996).

Such successes were despite all maintainers not having any previous experience of the use of graphics packages or spreadsheets. Competence in both of which is essential to unencumbered maintenance of Easy Speaker hierarchies, and later monitoring of user progress. In addition although they had some experience of Personal Computers for word processing and playing games they had little experience of the Microsoft Windows operating system. This meant that they were also unfamiliar with the Windows system applets that were utilised by Easy Speaker for retouching graphics, recording sound and editing scripts. Namely MS Paintbrush, Sound Recorder and Notepad. An additional hurdle was provided by the fact to access on-line help for Easy Speaker they needed to make use of MS Help, again an application they had never previously made use of.

With much trepidation all three maintainers managed to get to grips with generating Easy Speaker hierarchies. Initially all found the availability of the author helpful in clarifying aspects of Easy Speakers operation and entries in the user manuals and on-line help provided. However, this aid was more in the form of encouragement and enthusiasm than practical help. By the time maintainers completed the first QUIS (novice) they were tentatively editing their first scripts and all had constructed a couple of screens which contained links and speech.

As with any new piece of software maintainers found the process of learning how to use it quite daunting despite the authors aim to make the process as simple as possible. Initially when things went wrong they were more likely to blame to computer, or the software, rather than to pay heed to the error messages which signalled errors in their scripts, missing bitmaps, or unrecorded sound files. It was not a case of computer phobia but more one of cognitive dissonance.

When their subjective attitudes and opinions toward Easy Speaker were measured using the QUIS they rated it as being average. In terms of the adjectives QUIS uses they found it neither terrible nor wonderful, neither rigid nor flexible, and could not differentiate whether it had inadequate or adequate power. However they did rate it as being slightly frustrating and difficult. Despite discovering that to construct the visually appealing demonstration screens required quite a bit of initial effort, and could be frustrating, all rated the experience as being quite stimulating.

After around a week of experience the second QUIS revealed that attitudes had changed in a positive direction. So much so that ratings were significantly above the expected average rating. Whereas initially maintainers had found Easy Speaker to be neither terrible or wonderful they now rated it as being toward the wonderful side of the scale. From being neither rigid nor flexible, they now rated it as being more flexible than they thought. However the they still rated it as having about average power, or functionality from a maintainers point of view. Crucially whereas before they found it difficult they now found maintenance very easy. Frustration was gone and they rated maintenance as being quite satisfying. In sum they found the whole process to be even more stimulating than their novice ratings suggested.

As one might expect, through experience attitudes changed. Statistical analysis suggested this to be a very significant positive change. Although all maintainers were real, it cannot be assumed that maintainers in time pressured service environments will make similar progress. For example, if they initially have average, or lower, attitudes toward the maintenance of Easy Speaker hierarchies then they might potentially be put off from making full use of its capabilities. Worse still they may either use the hierarchies that came with the system without modification and lasting benefit to the user, or may simply cease to use it completely. Obviously this is the last thing that is wanted. In order to increase the appeal of Easy Speaker beyond the immediate visual impression what is needed is for novice maintainers to initial have a quite positive overall attitude toward it. This should not be so positive as to not allow for development, but should create an air of enthusiasm. Assuming that the three maintainers who participated in this short study are more-or-less representative, then it is useful to analyse their ratings beyond overall attitudes to consider specific strengths and weaknesses.

In terms of screen aspects novice maintainers rated Easy Speaker as being significantly better than expected. There were however specific features that were rated as being below average. The main area of concern was the amount of information that could be displayed on screen simultaneously. Further investigation revealed that this was a concern not only with the number of symbols that could be displayed, but also with the number of third party applets that contributed to screen crowding when editing scripts etc...

The first concern is perhaps more related to maintainers than to users. That is, maintainers were concerned that they could not display more than 19 symbols simultaneously on any one screen; eight standard symbols and eleven on the fast speech toolbar. Despite this concern it must be remembered that the more symbols that are displayed simultaneously the larger the cognitive overhead is likely to be placed on the potential user. Given a typical users cognitive deficits more symbols are likely to mean that they must resort to a scanning methodology to find the symbol they desire, rather than being likely to recall its location from memory. Increasing symbol density is perhaps not an option, as this leads to a decrease in symbol size which exaggerates any visual and cognitive deficits. The author suggests that the view that not enough symbols can be displayed is probably related to the fact that maintainers typically leaned toward a holistic approach. That is, their preference was include every conceivable symbol on a given screen rather than have to structure the hierarchy with intricate links due to reasons of time pressure. Alternatively they could have used fill-in phrases instead of individual words, e.g. "Please pass me _______". Both of which require more thought on the part of the maintainer.

On the second point, it must be remembered that Easy Speaker was designed to run in a screen resolution of 640x480 pixels. This is the lowest resolution offered by Windows and can lead to 'screen cram' when using more than one application simultaneously. The short term solution is to increase the screen resolution whilst editing, and then switch back to the standard resolution whilst users utilised Easy Speaker. For example a resolution of 1280 x 1024 gives a 427% increase in screen area, enabling four applications to be run simultaneously each with the equivalent of its own 640x480 pixel size window. With a 17" monitor this is perfectly feasible.

Concerns over the amount of information on the screen remained constant when the second QUIS was completed, and the maintainers considered expert.

A second area of concern was the ease with which a maintainer could go back to a previous screen when constructing and editing hierarchies. In general when testing the effectiveness of new links it is useful to be able to take those links as a user might and then to jump back to the screen the link was from. Obviously the same lack of reversibility, or undo, also relates to users. This function was not designed into the system as it was thought that it would simply place more demands on users. However in retrospect both a jump back to previous screen and undo speech symbol selection buttons are perhaps warranted as standard. In this respect these features will be implemented in the next version of Easy Speaker. They were left out as the hope was that maintainers would structure hierarchies in such a manner that links backward would be provided by the maintainer. However the ability to undo speech symbol selections is a glaring error on the part of the author pointed out by maintainers themselves.

All other aspects of the Easy Speaker maintainer and user interface were rated as average, or above average by novice maintainers.

When expert the only rating that remained below average was the concern over the amount of information on the screen. Only one rating dropped. Namely the whether the layout of the screens was helpful or not. However this only dropped very slightly to a fraction under what one might expect an average rating to be. In sum all other ratings remained constant, or rose above novice ratings. Overall expert ratings were very significantly above those expected.

Despite many ratings having improved, or remained constant there was no significant difference between novice and expert ratings overall. This suggests that experience is not related to how maintainers rate aspects of the Easy Speaker screen.

In terms of terminology and system information novice maintainers rated Easy Speaker as being significantly worse than expected. Areas of concern included task terms, terms on screen, messages which appear on screen etc... Only two areas from 17 concerning terminology and system information were rated was being average, or above average.

When expert however ratings rose overall to just over the expected average, although not significantly so. Many of the below average ratings of novice maintainers can be attributed to inexperience, as when expert their ratings rise to average, or above. It is useful therefore to consider those ratings that remained below average when expert, suggesting a design flaw rather than simply inexperience. One example of inexperience is in the mean response to "Performing an operation leads to a predictable result". Novice users mean response was 3.5, below average, whereas expert users responded with 6.5, above average.

In all there were six areas that remained below average. All but one rose between novice ratings and expert, but none was higher than the expected average rating of five. Interestingly all six of the *problem* areas concerned either error messages or feedback. The six areas were :-

Fig 6.9, Below average ratings for error messages and feedback items

QUIS item	Likert scale value 1 - 9	Novice mean	Expert mean
Error messages	unhelpful - helpful	3.00	4.00
Phrasing of error messages	unpleasant - pleasant	3.00	4.00
Error messages clarify the problem	never - always	2.00	4.00
Instructions for correcting errors	confusing - clear	3.00	4.00
Does the computer keep you informed about what it is doing	never - always	4.00	4.00
User can control amount of feedback	never - always	2.50	4.00

As can be seen from the above table expert ratings stabilised around 4.00, or just below average. Probably with more experience maintainers would have made better use of the error messages that were provided, as they would come to interpret their meaning and significance correctly. As a result they would take the appropriate action to correct the problem of their own volition. However this is unacceptable for novice users. What is needed is error messages that spell out the problem, listing probable causes, and corrective measures that might be taken. Perhaps what is needed is an approach similar to that taken by advanced GUI operating systems such as MS Windows 95.

When an error occurs the OS signals this to the user and then offers to either correct the problem itself, take the user through correcting the problems themselves, or offers further in-depth help detailing possible causes and corrective measures. Hopefully this approach will be followed in the next version of Easy Speaker.

As regards feedback, maintainers main complaint is that although they understood the simple scripts conversely they had trouble keeping track of them if they lost concentration during the development or modification stage. What was suggested is that a 'script debugging' window might be kept open while the maintainer emulated a real user and navigated the hierarchy. Thus when they clicked on any symbol or link the script the system was following could be seen simultaneously in a scrolling window. Again this would be an immense improvement, and will hopefully be included within the next version.

Perhaps more important than screen design, and terminology and system information is the ease with which Easy Speaker could be learned by maintainers. Learned in the respect of "Remembering specific rules about entering commands" for example. In addition "learning" also encompassed the content and amount of on-line help, other reference materials and tutorials.

The view of novice maintainers to learning to use Easy Speaker is particularly important. Encouragingly novice maintainers rated Easy Speaker as being of average difficulty to learn to maintain. This suggests that overall Easy Speaker was easy to learn to maintain as one might expect a lower than average rating given the maintainers lack of experience.

Although average overall initially maintainers rated 10 areas as being below average. These clustered around getting started with the system and learning of advanced features, the degree with which maintainers could explore features, and the amount and content of on and off-line help provided. However, as one might expect these initially below average ratings changed as maintainers gained more experience. Initially they found the maintenance procedures of Easy Speaker difficult to learn, but with more experience found this less so. Indeed when experienced they found Easy Speaker significantly easier than average to learn. However three areas from the original 10 either remained below average, or were rated as being worse suggesting a real difficulty.

Fig 6.10, The effect of experience on learning

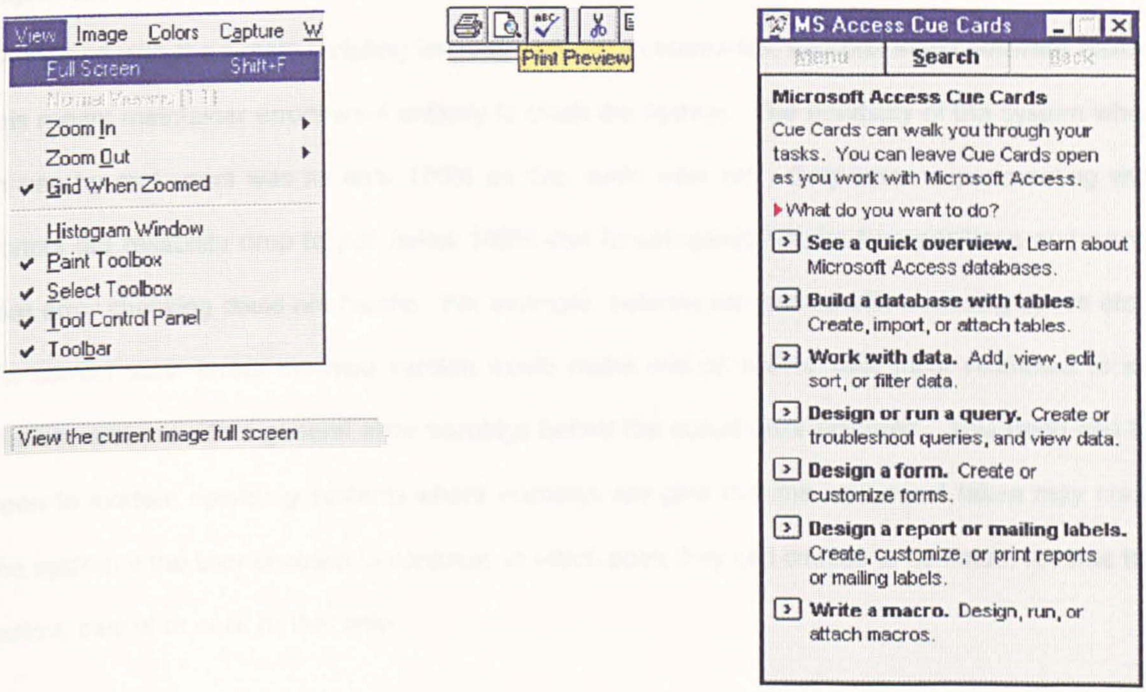
Item	Likert scale value 1 - 9	Novice Mean	Novice SD	Expert Mean	Expert SD
Learning to operate the system	difficult -easy	3.33	1.15	6.50	3.54
Getting started	difficult -easy	3.33	1.15	6.50	3.54
Learning advanced features	difficult -easy	3.50	0.71	9.00	-
Exploration of features by trial and error	discouraging - encouraging	4.67	0.58	6.00	-
Exploration of features	risky - safe	4.33	2.08	3.50	3.54
Help messages on screen	confusing - clear	4.50	2.12	4.00	-
Accessing help messages	difficult - easy	3.50	0.71	7.00	-
Content of help messages	confusing - clear	4.00	1.41	5.00	-
Amount of help	inadequate - adequate	3.50	0.71	4.00	-
Tutorials for beginners	confusing - clear	2.00	-	6.00	-

The first focused around the "Exploration of new features" and how risky or safe maintainers viewed exploration. Interestingly they rated exploration as being slightly more risky as they became more expert. Relevantly when expert they rated "Exploration of features by trial and error" as being encouraging with an above average rating of 6.0. This suggested that maintainers did explore features by trial and error as they found this to be encouraged, but found the process to be slightly risky. One can consider these views as being complementary as the more "things are played with, the more likely they are to break". This does not necessarily imply a design flaw as maintainers were willing to give exploration a try realising it might be risky.

The second area of concern remained "Help messages on screen" and the "Amount of help given". Further investigation revealed that these below average ratings did not apply to the on-line help guide, but to the help messages themselves and the depth into which they went. The question of functionality of help messages is really two fold. The first in the form of help in relation error messages has already being discussed. The second concerns constantly available help to maintainers. In this respect there was no constantly available help provided to guide maintainers, other than leaving the on-line help file open and in view on screen. In this respect the author hopes to provide in future versions a status bar that changes in relation to the maintenance task in hand. As a support MAC style "Bubble Help", or MS style "Tool Tips" may also be provided. In addition

MS style cue cards might be implemented so as to give semi-automated support to the completion of common tasks and act as a enhanced teaching metaphor.

Fig 6.11, A typical status bar, "Tool Tip", and "Cue Card"



Despite maintainers being concerned by the level of constant help provided they rated the paper based reference manual as being above average both novice and expert. The quality of the paper based manual may account in part for the significant improvement in ratings in learning between being novice and expert.

The final section of the QUIS concerned "System capabilities" and concerned aspects such as system speed, reliability, and ability to correct mistakes etc... Basically, the focus was on the "power" of Easy Speaker as a piece of software. Again encouragingly maintainers rated Easy Speaker as being around average in terms of system capability. Three areas were highlighted as being below average. Namely reliability, correcting mistakes, and whether or not novices can accomplish tasks using only a few commands.

Reliability was initially rated as being below average but rose to above average when rated by expert maintainers. This suggests that reliability tended to be in the "eye of the beholder" as hinted at earlier. Users are initially more likely to blame errors in scripts for example on the system rather than to accept the blame themselves. As Easy Speaker had been developed in such an object-orientated language as C/C++ this lead to high levels of system reliability and integrity. Combined with the simple scripting language with few commands, and prolonged software testing this meant maintainer errors were unlikely to crash the system. The reliability of the system when in use by real users was as near 100% as one could wish for. Only when experimenting with scripts did reliability drop to just below 100% due to unexpected script transpositions and errors that error checking could not handle. For example, entering file names with a leading space etc... To correct such errors the next version would make use of limited user input restricted to set choices and provide a general error warnings before the actual error occurred. This trend can be seen in modern operating systems where warnings are give that the action just taken may crash the system if the user chooses to continue; at which point they can choose to continue, reverse the action, cancel or seek further help.

Correction of mistakes and the ability to undo operations was also rated initially as being below average. However, as mentioned previously this was likely to be the combined result of inexperience, and the lack of an undo feature. Again the influence of inexperience is indicated as when expert maintainers rated the ability to correct mistakes and undo operations as being well above average. However, mistakes could only be undone by manually editing scripts etc... as there was no undo button that would return the system to its previous operational state. The feature to undo the immediately previous operation is planned for the next version of Easy Speaker as it would save a great deal of time and effort, and allow for more confident exploration of features.

Novice maintainers rated the ability to accomplish tasks knowing only a few commands as being below average. This suggested that maintainers must have a complete, or near complete, knowledge of the system and the scripting language before being able to make full use of it. Logic dictates that this is the case as one cannot make use of the system unless the simple scripting language is learned and the maintainer is fully conversant with it. As the scripting language only contained six or so essential commands, an no superfluous ones, the requirement is not as stringent as it may at first appear. However, future versions of Easy Speaker hope to move away, or at least hide, the scripting approach in favour of a more GUI approach. This would mean that maintainers could simply drop pictures onto the screen at the point where they wished them to appear, rather than have to learn a scripting language, regardless of its simplicity. This would obviously be of advantage for novice users.

When expert, maintainers rated the "system capabilities" of Easy Speaker as being very significantly above average. Furthermore, all their previous concerns detailed above changed to more favourable ratings, suggesting experience played a major role. This was confirmed by the very significant difference between novice and expert ratings. Only one are of concern was highlighted as being below average. Namely the ability of the system to warn the maintainer of potential problems. One solution as discussed is to enhance the error handling system.

SUMMARY

In sum the first version of Easy Speaker for Windows was considered as being above average as a piece of software. All maintainers managed to utilise the system to produce hierarchies for their intended *real* user(s), and had no real difficulty in doing so. Experience was shown to be the key to attitudes toward Easy Speaker and the maintainers ability to make constructive use of it. In many areas, as little as 10 hours experience made a significant positive difference when compared with when novice.

Despite successes areas for improvement were highlighted during QUIS analysis, and remedies were suggested for future versions. These centred around the need for more detailed error messages, help with errors, ability to undo and reverse actions. In short the highlighted concerns were of most relevance to novice maintainers. In future versions these concerns will be addressed in order to make Easy Speaker easier to be maintained by novice maintainers.

CHAPTER 7: Cognitive constraints on speech processing: Implications for the *spoken* output of ACD's

Many previous computer based Augmentative Communication Devices made available to the Learning Disabled language impaired population have translated some form of representational vocabulary into synthesised speech output. However the use of synthesised speech means that many systems produce speech which sounds unnatural, or "robot like", due to the approximation of limited grapheme-to-phoneme conversion algorithms employed.

Previous research has suggested that as speakers constantly monitor and re-analyse their own speech, in addition to decoding others', any increase in speech approximation would be likely to increase cognitive load and hence reduce possible remediative effects of listening to ones *own spoken* output.

The debate, centres around the need to switch to digitised speech systems, with many arguing that synthesised speech "remains adequate". It is suggested that synthesised speech is not adequate for the majority of the language impaired LD population, with digitised speech offering many advantages. It is further suggested that listeners to synthesised speech are under increased cognitive load, and that such overloading has clear negative implications for potential language remediation. In addition interaction with unimpaired listeners' is also likely to benefit from a less demanding form of output.

I hope to gauge the processing overhead incurred by synthesised speech by using a choice reaction time paradigm, which is suggested to assess cognitive load when obeying increasingly longer and more complex synthesised and digitised speech instructions.

INTRODUCTION

With the advent of new and ever more powerful computer based technologies research has blossomed in the area of computer based Augmentative Communication Devices (ACD) for the non-vocal and language impaired (Beukelman 1988, Alm, Arnott and Newell 1992, Pausch and Williams 1992). However, despite these in-roads many ACD's in widespread use tend to focus on the application of older technologies that have simply been reworked or revamped.

The continued use of synthesised speech output for the majority of ACD typifies the historically based approach. With the following quote typifying the attitude of commercial software developers, "Robotic text-to-speech is adequate for the handicapped, but will it have a place in offices ?" (Fay 1993). Fay goes on to champion the use of natural sounding digitised speech in office environments as a major technological breakthrough but precludes its use with those who might benefit most.

Synthesised speech has on the other hand consistently been shown to be more cognitively demanding to listen to and decode correctly, and is generally poorer in intelligibility (e.g. Clark 1983, Mirrenda and Beukelman 1990, and Mitchell and Adkins 1989). Combined with poor interface design and choice of symbol set, unoptimised output will logically impair the impaired user yet still further.

Unfortunately it is *usually* only the older technology based systems that ever find their way into any kind of respectable production figures and actually reach intended users in service professions such as the NHS. Such attention has tended to make many systems appealing, increased in perceived usefulness and applicable only to a small niche of the language impaired population. Unfortunately some systems of this genre are prone to place such a demand on the user that only the highly able and motivated are able to make use of their facilities.

As discussed in chapters two and three, many of these systems assume that users have near normal abilities in terms of both productive and receptive language understanding. More worryingly, assumptions are often made as to the users other cognitive abilities in terms of speed of basic processing, memory function, visuo-spatial awareness, and size of knowledge store etc... It is known however that productive and receptive language impairments generally mirror those in other cognitive modalities such as speed of basic processing, memory, attention etc... (Lally and Nettelbeck 1977, Newell and Walace 1978, Detterman et al 1992, Kail 1992).

Jensen (1979), Vernon (1987), Anderson (1990, 1992), Kail (1992), and Detterman et al (1992) have all for example demonstrated experimentally that speed of basic processing, measured via Inspection Time (IT), Choice Reaction Time (CRT) and other such performance measures correlate with IQ and to a greater extent Mental Age (MA). As discussed in chapter two, MA is generally below Chronological Age (CA) in those with impaired language functionality when measured using language independent tests such as Ravens progressive matrices, or the British Picture Vocabulary Scale for example. Impairments in memory subsystems are also common and have been reported by the majority of researchers in the area (Ellis 1970, Butterfield, Wambold and Belmont 1973, Anderson 1992, Detterman et al 1992, Kail 1992). Thus previous research confirms the notion of generally reduced cognitive abilities in the majority of those with language impairments and a LD.

Logically then, one should aim to produce an ACD which produces *spoken* output which incurs an acceptable cognitive load on decoding in relation to users', and indeed listeners', cognitive abilities. Indeed the role of ACD's as both communication system and as remediation tools hangs on this prerequisite.

Increased cognitive load can be detrimental as many studies of normal language processing have shown (Garrett 1984 and Dell 1986) we generally monitor and re-analyse our own speech output and modify it accordingly, Spoonerisms (Garrett 1976) being a good example. That is, we recognise speech errors as we monitor and re-analyse our spoken output rather than at the generative stage. Thus we are constantly under some degree of cognitive load even when speaking. If we have to cope with mentally planning utterances, driving an alternative production system, i.e. an ACD, and simultaneously listening to, and reprocessing its spoken output we are likely to begin to reach optimum processing levels, above which our performance is naturally prone to deterioration (Yerkes and Dodson 1908). Logically this will affect those with language impairments and poor general cognitive abilities disproportionately. Again, it makes little sense to overload the ACD user as the hope is that from such spoken output remediation will be a secondary benefit.

THE FORMS OF OUTPUT - SPEECH SYNTHESIS AND DIGITISED SPEECH

The notion of computer-based augmentative speech aids is relatively young, however despite its age many advances have been made as a consequence of the meteoric advancement of computing technologies. Whittington (1989) quotes that it is, "only within the last 10 years has *any* form of acceptable synthesised speech output from a computer reached the *masses.*"

During the 1980's the type of speech output was governed largely by machine requirements, that is in terms of the space required to store and process speech. This limiting factor imposed that a generative system must be used, that is speech synthesis, or as it is more commonly known "Text-to-Speech". Such systems were ideal as they required little in the way of storage and processing capacity and could generate any word or sentence, as and when required. The trade-off being one of quantity against quality (Yule 1986, Whittington 1989, Shneiderman 1992, Johnson 1992).

By their nature speech synthesisers produce only approximations of the fidelity of real human speech. The crudest synthesisers form each piece of speech by combing each of the phonemes which make up the pronunciation of a given word. They do so by converting stored graphemes to phonemes, or stored digital representations of each phoneme to be more precise.

Fig 7.1, A sample sentence translated into its phonetic script using phonetic transcription

Original

mary - had - a - little - lamb

Phonetic Script

mâr'ê - hàd - â - lît'l - làm

Pronunciation Key Click any pronunciation symbol or word to hear it spoken.

ă	pat	l	lid, needle	*th*	this
ā	pay	m	mum	ŭ	cut
âr	care	n	no, sudden	ûr	term, word, heard
ä	father	ng	thing		urge, firm
b	bib	ŏ	pot	v	valve
ch	church	ō	toe	w	with
d	deed, milled	ô	for, caught, paw,	y	yes
ĕ	pet		horrid, hoarse	z	zebra, xylem
ē	bee	oi	noise	zh	vision, garage
f	fife, phase, rough	o͝o	took	ə	about, item, edible,
g	gag	o͞o	boot		gallop, circus
h	hat	ou	out	ər	butter
hw	which	p	pop		
ĭ	pit	r	roar		**FOREIGN**
ī	pie, by	s	sauce	œ	*Fr.* feu, *Ger.* schön
îr	pier	sh	ship, dish	ü	*Fr.* tu, *Ger.* über
j	judge	t	tight, stopped	KH	*Scot.* loch, *Ger.* ich
k	kick, cat, pique	th	thin	N	*Fr.* bon

Primary stress ´bi•ol'o•gy **Secondary stress** ´bi'o•log'i•cal

From Microsoft Bookshelf 94

As one can appreciate these approximations when combined can sound fairly crude and can on occasion be barely understandable unless one devotes a large amount of cognitive effort to listening to them. Unlike digitised speech, synthesised speech *cannot* easily carry the intonnative qualities of natural speech which typically sets mood and tone during interaction. This in itself can further increase cognitive load as a mismatch between content and tone can occur frequently.

Combined with factors such as background noise which increase cognitive load further speech intelligibility is typically reduced dramatically (Fucci, Reynolds, Bettagere and Gonzales 1995). Bess and Humes (1995) for example have reported signal-to-noise ratios of between -6 dB to +6 dB in average classroom situations. At such levels speech synthesis intelligibility is likely to be severely affected. Other authors (e.g. Pratt 1987) have shown that natural speech is not significantly affected even above these levels.

More advanced synthesisers can, in ideal listening conditions, approach statistical equivalence with natural speech (Mirrenda and Beukelman 1987). With synthesisers such as DECTalk making use of full phonetic transcriptions, taking account of allophones etc..., for each word they *know* about, in addition to using standard generative methods. This is an improvement, as finer pronunciation features can be incorporated, but as with the truly generative methods the speech output still sounds somewhat "robot like" (Beukelman 1988, Srathmeyer 1990, and Shneiderman 1992). Recent researchers argue that synthesis based systems can only be improved by adding a vastly enlarged set of phonetic rules or by hand editing phonetic scripts (Karlsson and Neovius 1994). Unfortunately such improvements are only likely to be applicable in ideal listening conditions which seldom exist frequently in the real world.

Although many accept that synthesised speech is more demanding recent authors have also suggested that it affects memory for passages heard via synthesised speech. Higginbotham and Baird (1995) reported that miscomprehensions were common when subjects recalled simple passages. They suggested that these were more severe when they were originally learnt with a low quality synthesis system, as opposed to a high quality one. Furthermore they found that effects persisted even at slower rates of presentation, i.e. ≈ 5 wpm over the more normal ≈ 140 wpm. This suggests that some synthesis systems are even more demanding than others. Potentially this may mean that interactions between speaker and listener may be less memorable than if they had taken place in natural English.

Despite its short-comings speech synthesis enjoys a wide application and will undoubtedly continue to do so for the near future at least. In testimony to its strengths many augmentative speech aids, language readers and trainers make use of this system, with the key application being text-to-speech readers for the blind (Songco et al 1980). The Texas Instruments "Speak'n'Spell" being one of the more memorable of the consumerist speech synthesisers of the 1980's. Physics Professor Steven Hawking's synthesiser being one of the best non-commercial implementations.

Further testimony to the adherence to speech synthesis for augmentative systems is lent by Beukelman (1988, 1990) who reviewed no less than eight speech aids, all of which used speech synthesis. With Strathmeyer (1990) in a similar review achieving similar findings. Even today some of the latest research systems make use of speech synthesis, e.g. CHAT (Alm, Arnott and Newell 1992), CANDY (Pausch and Williams 1992), Deaf-Blind Sign-to-Speech (Shown on Tomorrow's World, May 5th 1993). However, it must be remembered that many of these systems are not intended for use by the LD population and have specific user groups for which they were designed. In the present review of currently available commercial devices in chapter three, around 75% or more made use of synthesised speech exclusively, and half of the remainder could make use of a combination of synthesised or digitised speech. With many of the systems sharing the same common underlying generative algorithm, differing only in terms of user interface, e.g. DECTalk being typical.

During the 1990's on the other hand computers' capacities are increasing at such a rate that Multimedia capable machines are appearing daily on the computer users desktop. These machines are capable of reproducing digital recordings of a normal speakers voice and thus have the potential to be used in augmentative communication systems, e.g. Lister (1992) and QED (1993).

Digitisation as opposed to synthesis offers a "tape recorder quality" which captures each acoustic aspect of real human speech, allowing for intonation, inflections, rhythm, gender identification etc... Despite its obvious advantages the use of digitised speech requires relatively extravagant hardware and software to make use of it to its full potential. Thus the use of digitised speech dramatically increases the complexity and length of software development and can and usually does preclude commercial interests. The implication being that digitised speech tends only to be developed for commercial interests while the language impaired are left the more easily implemented synthesis systems (Fay 1993).

One of the main arguments against the use of digitised speech initially was the storage requirements it needed, with short recordings typically placing a disproportionate demand on equipment of the time. With increasing power of readily available hardware this is becoming less of an issue. In addition, in a similar vain to the generative abilities of synthesised speech, digitised speech can also be used in a *generative* mode. This is achieved by using a 'dictionary look-up' method, whereby a text based sentence is scanned to separate out each word. Once the words have been isolated, their equivalent pre-recorded digitised word is 'looked-up' and appended to the current stage of digitised speech construction. When the digitised sentence is fully constructed the speech is output. Although this may sound time consuming with current generally available hardware this can be accomplished without the awareness of the user. The advantages of this method are clear in the respect that storage requirements are cut to a bare minimum, whilst retaining the human fidelity of the speech. Such systems employing this implementation include American telephone-based voice information and banking systems such as, Fidelity Automated Service Telephone (FAST), credit card information (Citibank Customer Service), airline schedules (American Airlines Dial-AA-Flight), where touch-tone keying results in appropriate information being vocally imparted (Shneiderman 1992). It is this very method that Easy Speaker uses to combat potentially massive storage requirements and redundancy whilst dramatically improving the quality of speech output.

In addition to which new compression techniques such as, Adaptive Differential Pulse Code Modulation (ADPCM), allow for a reduction in storage requirements of the order of 3:1 without any discernible loss of quality (Fay 1993). Later incarnations can offer even higher compression ratios with TrueSpeech offering around 20:1 compression with appropriate hardware assistance to give a quality very close to that of a good quality telephone line (DSP Group 1995). TrueSpeech can for example store 1 second of speech for every 1 kbyte of memory or disk space.

THE THEORETICAL CASE FOR USING DIGITISED SPEECH IN AUGMENTATIVE COMMUNICATION DEVICES FOR THE LANGUAGE IMPAIRED

Since its inception synthesised speech has been known to engage *different* cognitive resources in listeners as compared with natural speech. Many have assumed that these differences simply reflect a shift in emphasis rather than capacity. Historically however the intelligibility of speech synthesis has been found to be significantly poorer than natural speech (e.g. Clark 1983, Michell and Adkins 1989, and Mirenda and Beukelman 1990). Originally developed for military applications, speech synthesis has largely been superseded in its original role with the advent of more powerful hardware and software capable of utilising digitised speech.

Evidence from normal language users suggests that speakers monitor their own speech and re-analyse it in addition to decoding others' speech as the hear it (Marslen-Willson and Tyler 1980, Dodd and Cambell 1986, Rumelhart and McClelland 1986, Kintsch 1988, Bard, Shillcock and Altmann 1988, Ellis and Young 1988, Eysenck and Keane 1992). Thus normal speakers are under cognitive load both when speaking and listening, not to mention the load that is incurred as a result of planning the utterance to be spoken. There is no reason not to suppose that LD language impaired individuals follow this pattern. In terms of both communication and remediative possibilities a low demand, high quality form of output should be sought. Logically this should be digitised speech.

In more precise terms the much copied and rejigged "Cohort Model" originally proposed by Marslen-Wilson and Tyler (1980) proposes that the language processing system is flexibly structured so that various knowledge sources, e.g. syntactic, lexical and semantic, interact to produce an efficient analysis of spoken language. That is, word recognition is assumed to involve simultaneous bottom-up processes stemming directly from the spoken word and top-down processes based on the listener's expectations formed from contextual information (Frauenfelder and Tyler 1987). With processing occurring in parallel, such interactions seem highly plausible and appear to be supported by research data.

According to the "Cohort Model" language processing includes the following stages :-

1. Early in the auditory presentation of a word, those words known to the listener which conform to the sound sequence that has been heard so far will become active; this collection of potential candidates for the presented word is called the "Word-Initial Cohort"

2. Words belonging to this cohort are then eliminated either because they cease to match further information from the input word as this becomes available or because they are inconsistent with the semantic or other context.

3. Processing of the auditory word needs to continue only until the information available from the context and from the word itself is sufficient to eliminate all but one of the words in the "Word-Initial Cohort".

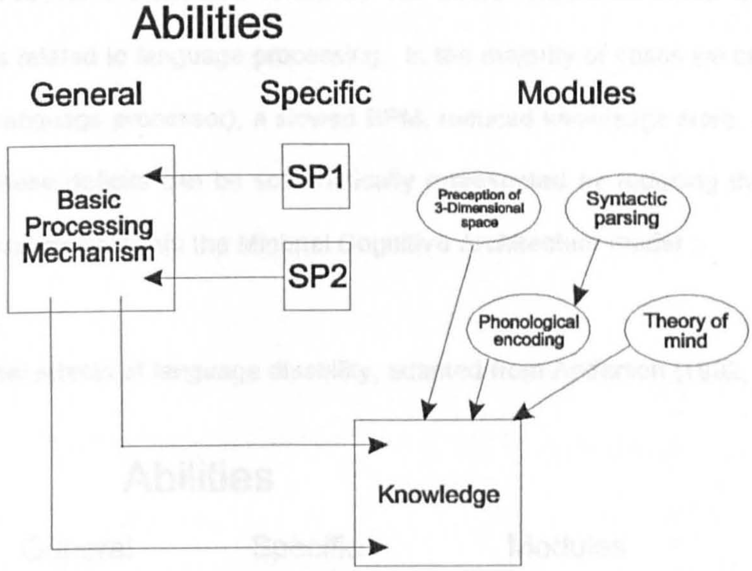
From Eysenck and Keane (1992)

As a direct result of parallel models of speech processing, such as the "Cohort Model", some authors have suggested that speech perception involves a specialised cognitive module or specific processor (Mattingly and Liberman 1988, Anderson 1988, 1990, 1992). Given the variability of speech and its transient nature, the existence of a separate processing module would seem plausible considering the speed and ease with which unimpaired language users decode speech (Darwin 1990).

However the majority of the language impaired population have below or well-below average productive and receptive language abilities. If we assume that they still continue to process speech according to the "Cohort Model" or another very similar system, then the fault probably lies with the module itself and/or the basic processing system which feeds it (Anderson 1992, Kail 1988, 1990, 1992, Klahr 1989, Tomporowski, Hyden and Applegate 1990, Detterman et al 1992).

If one takes Andersons (1992) "Minimal Cognitive Architecture" model or the functionally identical "Modal Model" of Detterman et al (1992) as a blueprint of the cognitive system one can envisage how this may occur :-

Fig 7.2, "The Minimal Cognitive Architecture", adapted from Anderson (1992)



Anderson's cognitive architecture is comprised of four main components, the Basic Processing Mechanism, the Specific Processors, the Modules and our store of Knowledge.

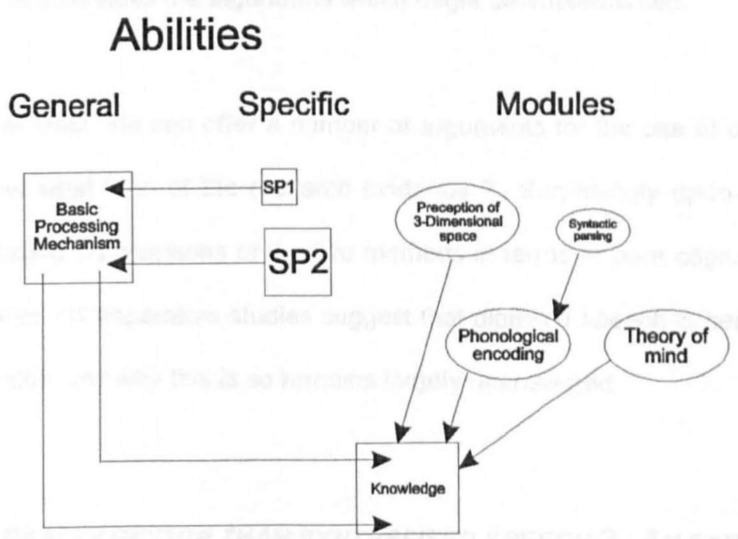
According to Anderson the Basic Processing Mechanism can be regarded as the "Homunculus" which governs the whole system, feeding to and collating information from each sub-system. With the speed of the Basic Processing Mechanism (BPM) being likely to determine how effectively the system operates as a whole (Anderson et al).

According to Anderson's (1992) differentiation hypothesis this effectively means that the slower the Basic Processing Mechanism (BPM) and the poorer the latent ability of a Specific Processor (SP) the worse the overall performance. With this theory applying to both expressive and receptive language abilities.

To use a common analogy the BPM can be likened to a computer CPU, with the SP, in this case concerned with speech perception, being the software algorithms that deal with the actual decoding processes. With Anderson et al (1992) noting that lower at levels of intelligence small increases in BPM speed show largest gains in terms of latent abilities of SP's.

Thus as it is uncommon to find language impairment, without other cognitive deficits (Waltz and Pollack 1985 and Merrill and Jackson 1992), we can utilise Andersons model to account for deficits in other areas related to language processing. In the majority of cases we can assume a deficit in SP1 (main language processor), a slowed BPM, reduced knowledge store, and general module problems. These deficits can be schematically represented by reducing the surface area of each of these components within the Minimal Cognitive Architecture model :-

Fig 7.3, Theoretical effects of language disability, adapted from Anderson (1992, 1993)



*NB - As represented above by reduced area of the BPM, SP1, Knowledge, and Syntactic Parsing

In relation to the current hypothesis[1] both the Cohort Model (Marslen-Wilson and Tyler 1980) and the Minimal Cognitive Architecture model (Anderson 1992) have clear implications for the type of *speech* an ACD should theoretically produce, namely digitised.

In terms of the Cohort Model, digitised speech should allow for 'on-the-fly' decoding, for nominal cognitive effort, as compared with synthesised speech, which by virtue of being an approximation to natural speech should require more processing. Indeed this is an inevitable consequence of

1 Synthesised speech will place the listener and/or *speaker* under greater cognitive load as compared with digitised speech.

approximation to natural speech should require more processing. Indeed this is an inevitable consequence of approximation to the norm and is probably still present to some degree regardless of learning effects. Although output from Easy Speaker can consist of concatenated digitised speech this is still in many cases more natural than that offered by many synthesis systems especially for the user group in question.

The Cohort Model meshes nicely with Anderson's' concepts of the Basic Processing Mechanism, and Specific Processors which suggests for such an algorithm to be implemented by a language impaired individual synthesised speech would take longer to decode, using up more of an already limited cognitive resource. That is, Anderson's model provides the cognitive hardware whilst Marslen-Willson et al's provides the algorithms which might be implemented.

Theoretically then at least one can offer a number of arguments for the use of digitised over synthesised speech, but what then of the research evidence ? Surprisingly up-to-press there have been few well-controlled comparisons of the two methods in terms of pure cognitive load each incurs within the listener. Comparative studies suggest that digitised speech is *better*, but the question of how much better and why this is so remains largely unanswered.

IS DIGITISED REALLY BETTER THAN SYNTHESISED SPEECH ?: AN EXPERIMENTAL EVALUATION

The Paradigm

Only recently have high quality studies been conducted in to the overhead imposed by synthesised speech e.g. Fucci et al 1995, Higginbotham 1995. These generally have concentrated on listeners perceptions and the intelligibility of it. However there is little research which concentrates on the cognitive overhead imposed by synthesised speech which propose also a model to account for such anomalies, other than pointing toward the lacking acoustic qualities of machine generated speech, and the implication for recognition.

Formally assessing whether digitised speech actually does impose a greater cognitive load upon listeners and in this case *speakers* is fraught with difficulty. One must indirectly measure, and therefore infer, the degree of cognitive load a listener/speaker is under by utilising some simple measure of performance. With this in mind I have opted to utilise a Choice Reaction Time (CRT) paradigm in order to assess the cognitive load incurred when listening to and decoding either synthesised or digitised speech. Using such a measure is appealing as Brewer 1980, Brewer and Nettelbeck 1979, Kirby et al 1982 and Wade et al 1978 have found good correlation between CRT, IQ and MA. In addition both Hick (1952) and Hyman (1953) have noted that RT increases linearly with the $\log_2$ of the number of response alternatives. This is highly relevant as one would propose that synthesised speech requires the speech to be broken down into more units which take progressively longer to process due to its approximation. Thus word and ultimately sentence recognition should take progressively longer as compared with that required for digitised speech as length and complexity increase.

The actual paradigm being followed is a modified version of Jensens (1980, 1982, 1987b) methodology for assessing processing speed in relation to intelligence. Under this modified paradigm subjects are required to press a given coloured key on the basis of task instructions given in either synthesised or digitised speech. With these *spoken* task instructions increasing in length and complexity over trials so as to increase cognitive load (Merrill and Jackson 1992), and hence the ease with which the *spoken* instruction can be decoded and acted upon. Both presentation and measurement are handled by computer to ensure a very high degree of accuracy (+/- 1 ms).

Design

A fully systematic repeated measures design was employed as each subject would be tested with both types of task instruction (synthesised and digitised speech) and their resulting CRT's measured. The same tasks would be used in both conditions.

Subjects

Subjects consisted of a heterogeneous group of 20 psychology students who were all in their 20's, with the ratio of males to females being approximately equal. Subjects were assumed to be of above average intelligence and have fully developed language usage and a *normal* speed Basic Processing Mechanism. All subjects had normal vision and hearing and none were colour blind. All were unpaid volunteers who were selected at random from an available population.

Justification for the use of unimpaired subjects

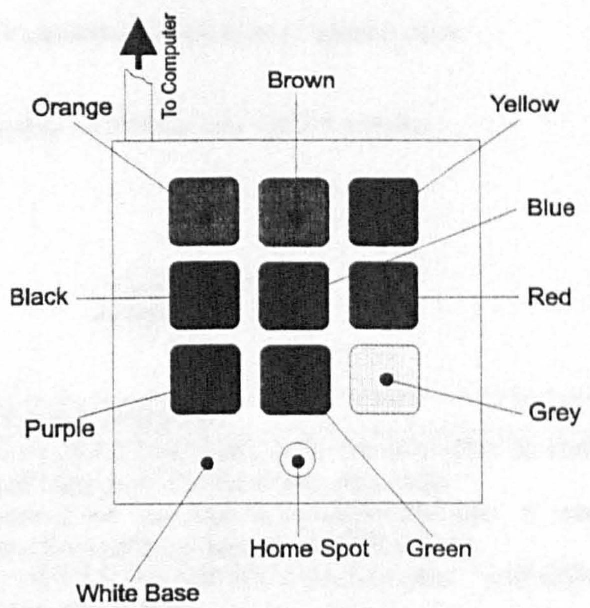
Although the intended users of most ACD's are unlikely to possess the same levels of cognitive and other abilities as unimpaired individuals, they can in certain areas be regarded as utilising similar resources but at lower functional thresholds. Such a notion is central to the theories of Anderson (1992), Nettelbeck (1985) , Strollery, Rabbitt and Moore (1990), Merrill and Jackson (1992), Rabbitt and Maylor (1991), Detterman et al (1992), Kail (1992) and others' who attempt to evaluate and explain cognitive functioning by comparing normal individuals with those considered to be impaired. Strolley, Rabbitt and Moore (1990) for example using 2, 3 , 4, and 5 Choice Reaction Time paradigms have consistently found that lower IQ Test Scores (IQTS) are mirrored by slower CRTs. Strolley et al have also reached similar conclusions when considering age, the ingestion of alcohol and more relevantly verbal ability. In terms of the current paradigm one would justify the use of normal subjects by suggesting that any differences would be likely to be mirrored and probably exaggerated in those lower in cognitive and language ability. Indeed Anderson (1992), Kail (1992), and Detterman (1992) addressing this issue in Inspection Time (IT) and related paradigms found that in addition to responses times increasing as Mental Age (MA) decreased, variability in responses also increased.

Using subjects of normal abilities also meant that tasks were more likely to be understood correctly, that more complex tasks were able to be utilised and that responses showed less variability and hence were more stable. Thus the use of normal subjects in assessing cognitive load is more likely to have provided a less contaminated picture of the processing overhead incurred when decoding synthesised speech. That is, variables such as hearing ability, knowledge for colours, visuo-spatial knowledge etc... remained more-or-less constant across subjects; whereas these variables would have been likely to unduly effect CRTs if language impaired subjects LD had been used exclusively.

Apparatus

A 9 key, colour coded keypad was constructed in order to measure Choice Reaction Times (CRTs) of subjects in response to synthesised or digitised task instructions. All colours were randomly selected from an available population of non-similar colours and were randomly assigned to each key. The actual box housing the keypad was white (a colour not used for keys) and was clearly marked with a 'home' spot from where each subjects pressing finger would be placed at the start of each trial :-

Fig 7.4, The 9 key colour coded keypad used for CRT measurement



This keypad was linked to a 20 Mhz Elonex 386 portable running a bespoke program written by the author which was used for both presentation and measurement. The program written in Borland C/C++ could reproduce both synthesised and digitised speech which had previously been entered as ASCII text files or digitised using another machine.

Synthesised speech was generated using an industry standard piece of text-to-speech software from First-Byte Inc. which is considered to be representative of synthesis systems on offer, and comparable with the popular DECTalk synthesiser. Whereas digitised speech was recorded as complete phrases using a 33 Mhz Apricot LS Pro 386 desktop with a 16 bit stereo sound card capable of CD quality. However in this case all samples were digitised at 11,000 Hz using mono 8 bit Pulse Code Modulation (PCM). Digitised speech was reproduced using a C/C++ library developed by the author for use with Easy Speaker.

Both synthesised and digitised speech were output via a Covox Inc. Voice Master Key II digitiser attached to the parallel port of the Elonex portable. This *speech* output was then amplified using a pair of Ross portable speakers. Volume levels remained constant throughout.

Both synthesised and digitised speech tasks were identical in content and were of approximately equal length. The table below shows the speech examples and CRT tasks given to each subject along with the length in seconds of each type of speech used.

Fig 7.5, Speech examples, CRT tasks and speech lengths

Examples and sample CRT task	Synthesised Speech Length (secs)	Digitised Speech Length (secs)
Short Example: "This is an example"	1.19	1.06
Long Example: "Mary had a little lamb, its fleece was white as snow; And every where that Mary went the lamb was sure to go. It followed her to school one day, that was against the rule; It made the other children laugh and play to see the lamb at school. And so the teacher turned it out, But still it lingered near; And waited patiently about till Mary did appear. Why does the lamb love Mary so the eager children cry; Why, Mary loves the lamb, you know, the teacher did reply."	28.00	27.88
Task Example: "Press the blue key"	1.50	1.54
Mean length	10.23	10.16
SD	15.39	15.35

Experimental CRT tasks	Synthesised Speech Length (secs)	Digitised Speech Length (secs)
Task 1: "Press the black key"	1.56	1.38
Task 2: "Press the key to the left of the grey key"	2.68	2.63
Task 3: "Press the key above the black key"	2.22	2.09
Task 4: "Press the key which is enclosed by the brown, blue and red keys"	4.13	4.47
Task 5: "Press the key which is above the red key but to the right of the black key"	4.40	4.19
Task 6: "Press the key which is in the second column and third row"	3.85	3.56
Trail 7: "Press the key which completes the L shape which begins with the brown key"	4.81	4.19
Trail 8: "Press the key which is below the middle row and diagonally in line with the red key"	5.38	5.69
Task 9: "Press the key which is to the right of the black key and diagonally in line with the brown key and diagonally in line with the green key"	8.06	8.28
Mean length	4.12	4.05
SD	1.94	2.06

Earlier piloting with the experimental CRT tasks had suggested that the tasks chosen would be likely to illicit the desired increases in cognitive load in normal subjects.

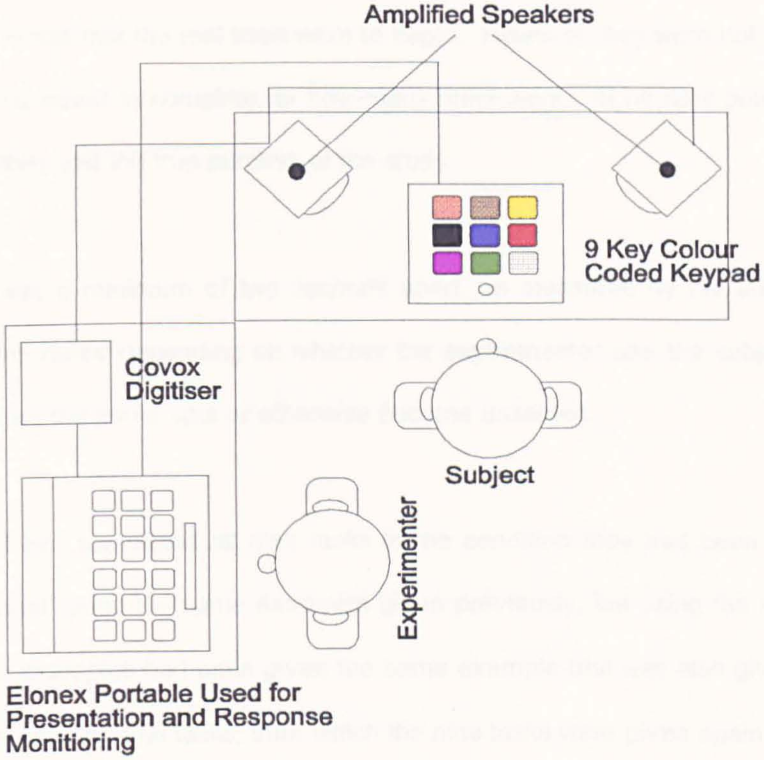
Procedure

Subjects were assigned to conditions on the basis of a ABBA randomisation scheme. For example :-

> S1 (synthesised-digitised), S2 (digitised-synthesised), S3 (synthesised-digitised) .. S20 (digitised-synthesised)

Once subjects had been assigned to the given condition they were led to the room where the experiment would take place. Apparatus was laid out as in Fig 7.6.

Fig 7.6, Apparatus arrangement and position of experimenter and subject



Once each subject was seated comfortably they were given the standardised instructions shown in Fig 7.7., shown the 'home-spot', briefed as to its purpose, and then asked if they understood what they were to do.

When testing each subject the following standardised instructions were given :-

Fig 7.7, Standardised instructions

1. Your task is to press the correct key on the key pad when told to verbally by the computer. Do this as quickly as you can after the instruction has finished, but you must make sure the instruction is complete, don't try to guess and don't move your finger from the home spot until you are sure which key to press. If you don't know which key to press, press the key you think you should. When you know which key to press, press only one and when you press it, just tap the key once quickly.
2. Do you understand what you have to do ? Can you see all the colours on the key-pad ?

Each subject was then given a short example of how the speech would sound - "This is an example". All examples and tasks were either given in synthesised or digitised speech to match the condition the subject had been assigned to. Each subject was then given a longer example -

"Mary had a little lamb ...". Once the subject had passively listened to both examples they were asked to complete a sample trial, "Press the blue key". Once the example trial had been completed subject were told that the real trials were to begin. However they were not told what type of tasks they would be asked to complete, or how many there were. At no time before or during the experiment were they told the true purpose of the study.

Each trial (task) was a minimum of two seconds apart, as measured by the administration program, but could be varied depending on whether the experimenter saw the subject remove their 'pressing finger' from the home spot or otherwise become unsettled.

When the subject had completed all nine tasks in the condition they had been assigned to first they were 'reoriented' using the same examples given previously, but using the opposing style of speech. After the examples had been given the same example trial was also given. The subject was then told to expect the real tasks, after which the nine tasks were given again. At no time was the subject told to expect the same nine tasks in their second condition as they had received in their first.

When the subject had completed both conditions they were debriefed as to the real purpose of the experiment and were shown a printout of their results. Each subjects results were then given a sequential subject number, i.e. S1..S20 according to the condition they had completed first.

RESULTS

Descriptive statistical analysis suggested that there was a difference between valid Choice Reaction Times (CRTs) produced in response to task instructions given in synthesised and digitised speech; synthesised speech (M (136) = 2.017 secs, SD = 2.214) : digitised speech (M (145) = 1.626 secs, SD = 1.413). Valid CRT's were CRT's where the correct key was pressed in response to a *spoken* task instruction.

Fig 7.8, Descriptive statistics for synthesised and digitised speech tasks

	Synthesised Speech	Digitised Speech
Mean (seconds)	2.02	1.63
SD	2.21	1.41
Variance	4.90	2.00
Valid CRT's	136	145
Invalid CRT's	44	35
Total Trials	180	180

In order to evaluate whether this was a significant difference a dependent t-test was applied (t (120) = -2.13, p. = .035, Sig. .05 level).

Although there was a significant difference between synthesised and digitised task instruction, an independent t-test was carried out to assess whether this could have been produced by a difference between the lengths of task instruction for each type of speech. The resulting t-value (t (16) = 0.08, p. = .941, NS) proved to be non-significant. Confirmation was provided by a Spearman rank correlation coefficient (r_S (9) = .933, p. < .001, Sig. .001 level) which proved to be very highly significant.

A further independent t-test was carried out to assess whether differences in the length of examples and the example task could have influenced later experimental task performance. Again the resulting t-value (t (4) = 0.01, p. = .996, NS) proved to be non-significant. Confirmation was provided by a Spearman rank correlation coefficient (r_S = 1.00, p. < .001, Sig. .001 level) which proved to be very highly significant.

Given that there is a non-significant difference between the length of examples, the example task and experimental task instructions for each type of speech, one can conclude that differences in CRT's were likely to have been produced as a result of the type of speech used in each task instruction.

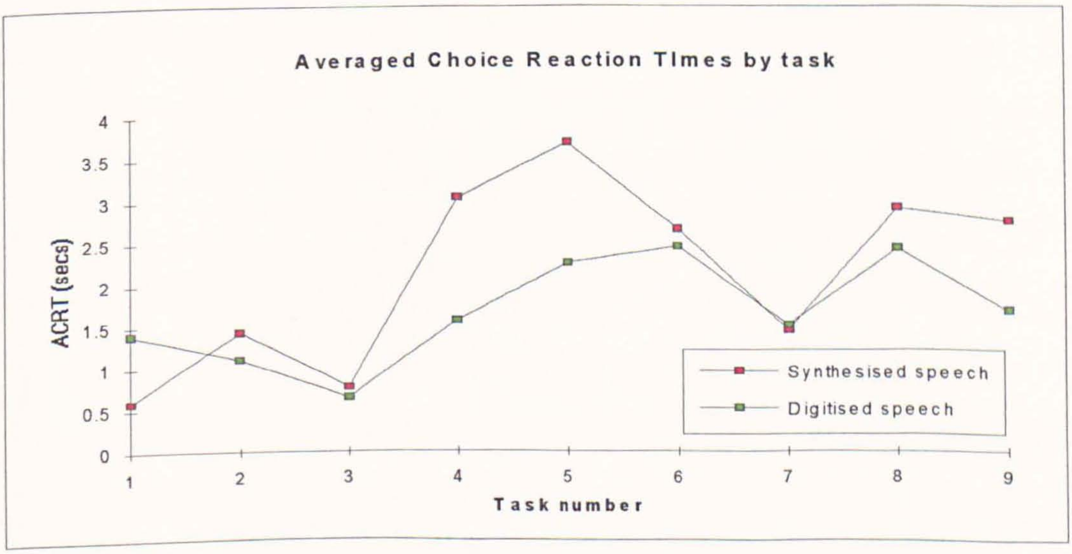
In order to more clearly illustrate these differences Averaged Choice Reaction Times (ACRT's) were computed using a similar method to that for Averaged Evoked Potentials (AEP's) in EEG studies; i.e. CRT's for a given task were totalled across subjects and then divided by the number of valid CRT's for that task :-

Fig. 7.9, Averaged Choice Reaction Time's (ACRT's) by task

Task Number	Synthesised Speech Mean (secs)	Digitised Speech Mean (secs)
t1	0.597	1.400
t2	1.433	1.095
t3	0.773	0.652
t4	3.038	1.568
t5	3.684	2.264
t6	2.663	2.457
t7	1.461	1.525
t8	2.914	2.435
t9	2.764	1.695

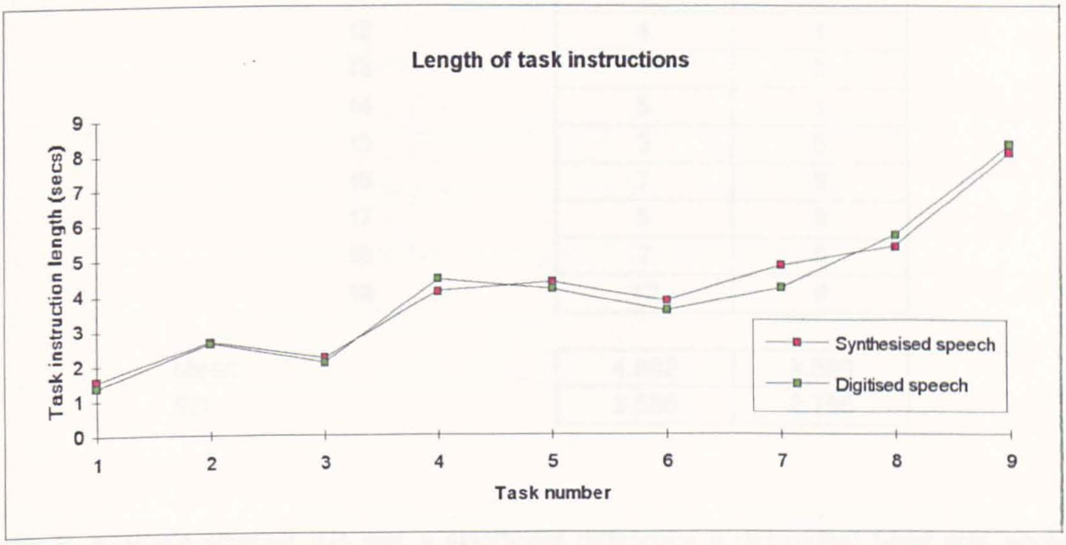
Using ACRT's in this way enables smoothed curves to be plotted, which can then be compared with a theoretical performance curve based on the length and complexity of each task :-

Fig. 7.10, Plot of Averaged Choice Reaction Time's (ACRT's) by task



The graph below represents the theoretical curve based on the length and complexity of each of the task instructions :-

Fig. 7.11, Plot of curve that show increases in ACRT's based on length of each task instruction in each condition



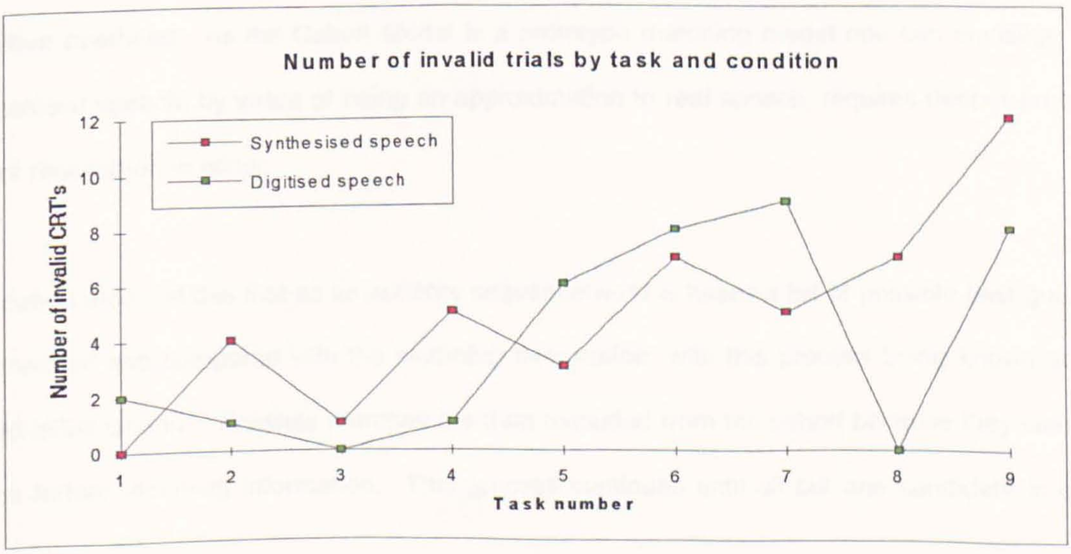
In addition to examining CRT performance CRT error rates were also measured, that is, tasks that were invalid because the incorrect key was pressed in response to a *spoken* task instruction. Descriptive statistical analysis again suggested there ware a difference in the number of invalid trials per task for task instructions given in synthesised and digitised speech; synthesised speech (M (45) = 4.889 secs, SD = 3.568) : digitised speech (M (35) = 3.889 secs, SD = 2.790).

Fig. 7.12, Number of invalid trials by task

CRT Task	Synthesised invalid Trials	Digitised invalid Trials
t1	0	2
t2	4	1
t3	1	0
t4	5	1
t5	3	6
t6	7	8
t7	5	9
t8	7	0
t9	12	8
Mean	4.889	3.889
SD	3.586	2.790

In order to evaluate whether this was a significant difference a dependent t-test was applied ($t_{(179)} = -1.45$, $p. = .150$, NS) which is non-significant. However despite this non-significant difference in invalid trials it is relevant to 'eyeball' a plot of the data as one can clearly see how the number of invalid trials in the synthesised speech condition increases in an almost linear fashion as the task length and complexity increases. Whereas increases for digitised speech are more erratic :-

Fig 7.13, Plot of the number of invalid trials by task and condition



This may imply that synthesised speech incurred a consistent overhead whilst the digitised speech showed more range due to its naturalness.

SUMMARY OF FINDINGS

Statistical treatment of results suggests that synthesised speech incurs a significantly ($p. < .05$) greater cognitive load even for subjects of normal language and cognitive ability when compared with high quality digitised speech. Error rates although not significantly different ($p. = .150$) do suggest that synthesised speech in addition to requiring more processing effort is also likely to induce a greater number of decoding errors. Synthesised speech produced 45 (25.0%) invalid trials out of 180, against 35 (19.4%) with digitised speech. These findings are lent further credence as there was no significant difference between the lengths experimental tasks for each type of speech ($p. = .941$). Indeed the lengths were so similar as to be correlated to a high degree ($p. < .001$). Thus it seems unlikely that CRT differences between the types of speech were an artefact of the difference in lengths of each task instruction.

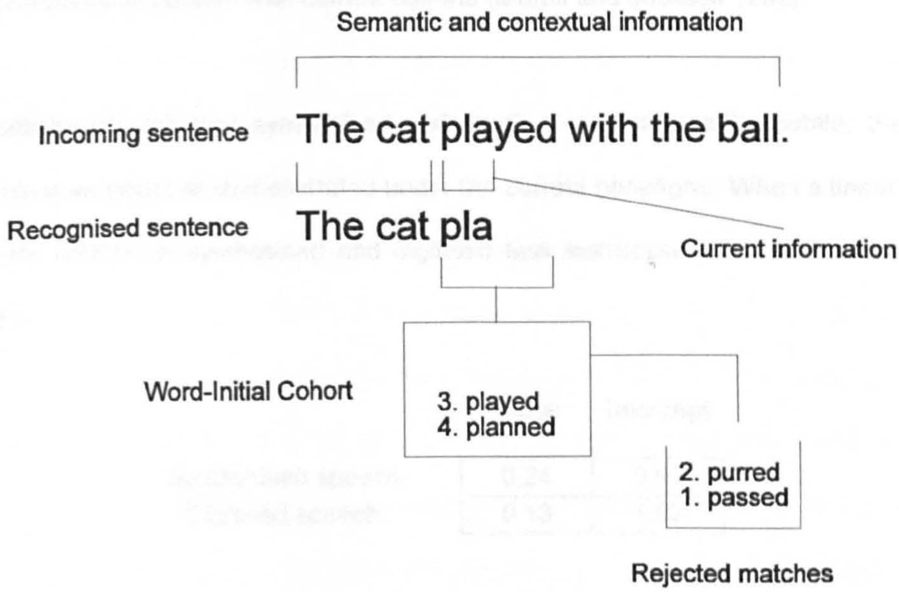
DISCUSSION

As synthesised task instructions incurred a significantly larger cognitive overhead as compared with digitised instruction the hypothesis that high quality digitised speech should be used in augmentative systems such as Easy Speaker is supported.

This finding lends support for the application of the Cohort Model of Marslen-Wilson and Tyler (1980) and other generic models in understanding how synthesised speech might incur a larger cognitive overhead. As the Cohort Model is a prototype matching model one can envisage how synthesised speech, by virtue of being an approximation to real speech, requires deeper processing for recognition to occur.

The cohort model states that as an auditory sequence/word is heard a list of possible best-guesses is generated and compared with the incoming information; with this process being known as the "Word-Initial Cohort". Possible matches are then excluded from the cohort because they cease to match further incoming information. This process continues until all but one candidate is eliminated from the "Word-Initial Cohort". However recognition is not based solely on gross physical features of the speech but also on semantic, alliterative and other contextual information which is taken into account during generation of each cohort and during elimination of poor candidates. That is, bottom-up and top-down processing strategies are utilised in parallel.

Fig. 7.14, The Cohort Model in action



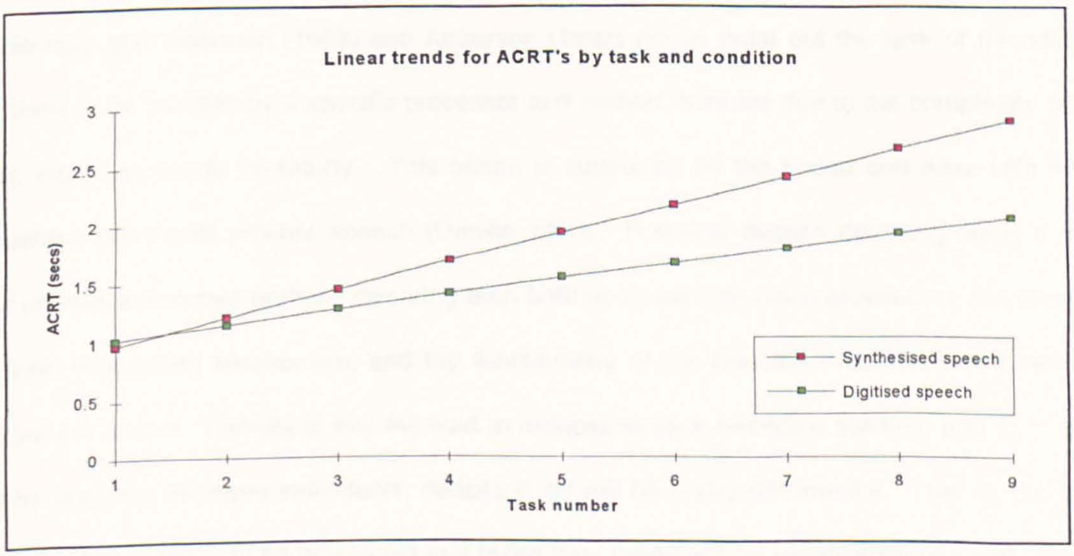
By using the Cohort Model to construct the representation above one can appreciate how each word might be recognised on the basis of its gross physical characteristics, i.e. "pla", and upon the context in which the word appears, i.e. cats play with balls. However, if the above example were spoken in synthesised speech as opposed to digitised human speech, it is likely that we would need to process the incoming information more deeply in order to find an exact match for each word. More precisely due to the approximation of synthesised speech a larger Word-Initial Cohort might need to be constructed, which then might not be reduced as quickly as normal, and hence the next piece of incoming information might overlay that being currently processed resulting in reduced efficiency. Thus the BPM and SP of Anderson et al are placed under a greater load when implementing a recognition algorithm such as the Cohort Model. As these are already thought to have less capacity in the LD population the implication is clear.

In support of the notion of information overlay Bard, Shillcock and Altman (1988) found that recognition under less favourable conditions was considerably slowed and that listeners tended to make use of subsequent rather preceding context in order to identify words to which they were listening. This suggests that listeners were waiting for more contextual and other information to become available so as to correctly identify words using a scheme such as that proposed by the Cohort Model. Thus, processing demands are likely to increase dramatically as previous cohorts are still being processed in parallel with current cohorts (Merrill and Jackson 1992).

Although differences between synthesised and digitised speech can be subtle, they are large enough to have an effect as demonstrated under the current paradigm. When a linear trend plot is generated for ACRTs of synthesised and digitised task instruction these subtle differences are highlighted :-

	Slope	Intercept
Synthesised speech	0.24	0.97
Digitised speech	0.13	1.02

Fig. 7.15, Cognitive load incurred as a result of synthesised and digitised task instruction



With cognitive load based on the calculated linear trend for ACRT's mirroring the increase in length and complexity of each task instruction. The subtle differences between synthesised and digitised speech are highlighted for the longest and most complicated tasks as a result of increased processing demands resulting in reduced efficiency.

As can be clearly seen from the linear trend plot above the more complex the task the larger the cognitive overhead appears to be in decoding. The most simple tasks have little effect, and the most rudimentary have none.

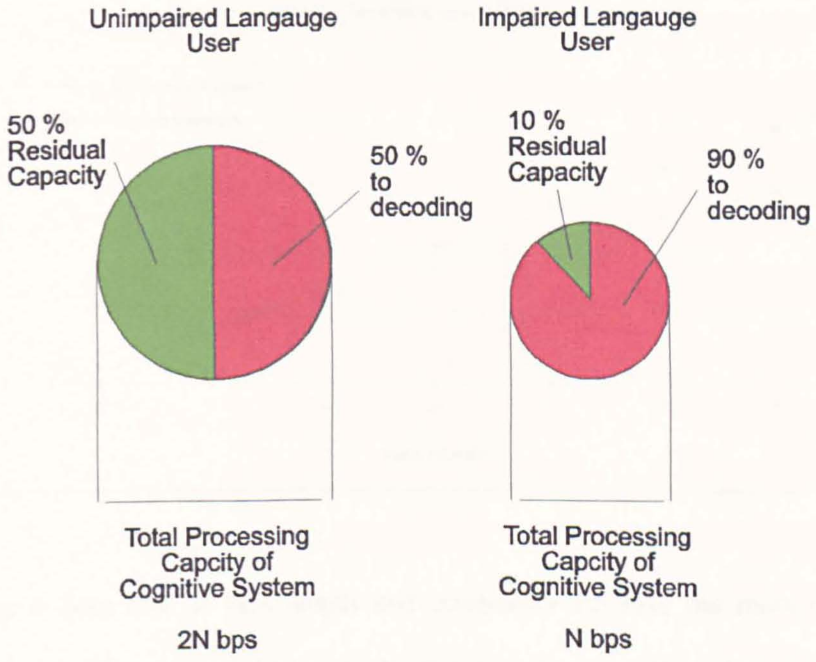
This linear increase is best explained by Shannon and Weaver et al (1949) who stated that for a given reduction in information processing rate the absolute increase in decision times will increase linearly as a function of the task information processing load. In this case more bits of information need deeper processing the more complex the message due to the approximation of synthesised speech. For example, if an unimpaired subject can process 2N bits of information per second, then the more bits, the greater the cognitive load (Rabbitt and Maylor 1991). In this case more bits of information are generated due to the probable way we decode synthesised speech as outlined above. For example, a synthesised speech word might generate and require the processing of 7 bits of information whilst the same digitised word might require 5 bits. This translates to 7/2N bits per second (bps) and 5/2N bps respectively.

As Mattingly and Liberman (1998) and Anderson (1992) rightly point out the task of decoding is very likely to be handled by a specific processor and related modules due to the complexity of decoding incoming words 'on-the-fly'. This notion is supported by the speed and ease with which unimpaired individuals process speech (Darwin 1990). However despite decoding being a massively parallel interactive process involving both bottom-up and top-down processing, the speed of the Basic Processing Mechanism, and the functionality of the Specific Processors and Modules involved are crucial. Deficits in any will lead to reduced speech decoding abilities, and as is more likely in language impaired individuals, deficits in all will be highly detrimental. That is, the number of bits that are able to be processed in a given time frame will be significantly reduced. As it is postulated that synthesised speech generates and requires the processing of more bits of information, one would logically want to avoid the use of such speech opting instead for more natural digitised speech which places less of a burden on the cognitive system outlined.

In more concrete terms an unimpaired language user might process 2N bits of information per second whilst a language impaired individual might process only N bps. Thus using the example above, a synthesised word might incur a processing load 7/2N bps for a unimpaired individual versus 7/N bps for an impaired individual.

This can be graphically illustrated if we consider this processing capacity as a 'processing pie' which is divided appropriately given the tasks in hand (Shiffrin and Dunmis 1981). With the size of the overall pie being finite and potentially static (Case 1985) :-

Fig 7.16, Total cognitive processing capacity and its relation to language processing

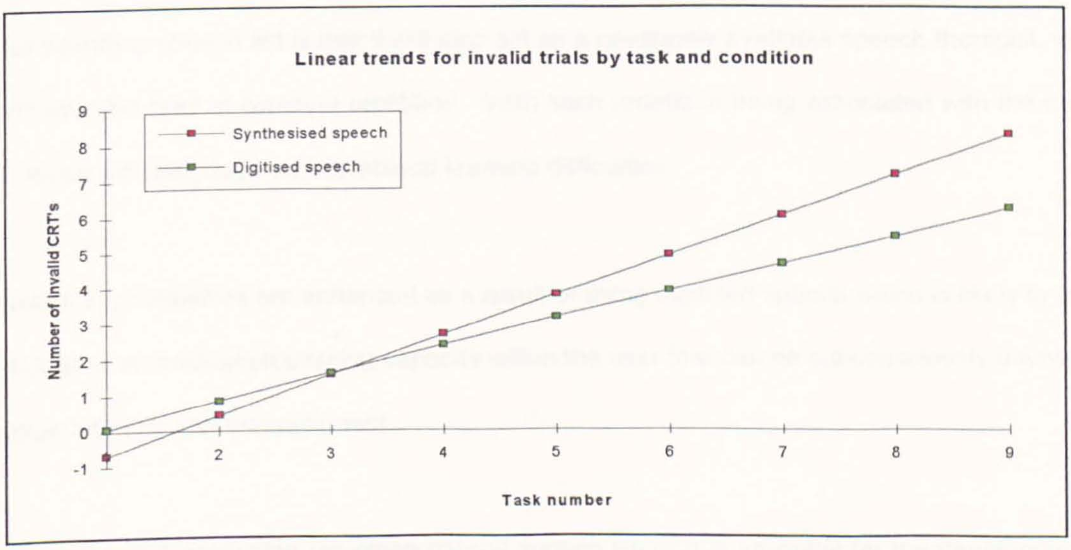


In the scenario illustrated above an unimpaired language user need only devote 50% of their total processing capacity to decoding of speech, whilst an impaired language user must devote 90% of their capacity to maintain an equivalent level of decoding power (indicated by area of the red segment), Merrill and Jackson (1982). Given the high demands of speech processing, the use of synthesised speech may exhaust the capacity of the language impaired individual resulting in decreased efficiency and even decoding failure. It is to this potential for decoding failure that I now turn.

If one examines error rates (invalid trials), for each type of speech, although not significantly different (p = .15) synthesised speech does appear to induce slightly more decoding errors (6.6 %) over digitised speech. This can be illustrated best when data for invalid trials are plotted as a linear trend :-

	Slope	Intercept
Synthesised speech	1.12	-0.69
Digitised speech	0.77	0.06

Fig 7.17, Linear trend plot of synthesised and digitised invalid trials



It is interesting to note that as task length and complexity increase the more decoding errors (invalid trials) occur. This can be taken as supportive of earlier assertions that synthesised speech is more cognitively demanding, as one would propose that the greater the demand placed on finite cognitive resources the greater the probability of decoding errors occurring. Indeed this assertion seems to be supported in terms of ACRT's and error rates, Fig 7.15 and 7.17 respectively.

One would also propose that as total processing capacity decreases as a result of language and other cognitive impairments in addition to CRT's increasing so too would the number of decoding errors.

In sum, I have suggested both from a theoretical and experimental standpoint the justification for the use of digitised speech over synthesised in ACD's. That is, digitised speech is least cognitively demanding in terms of decoding effort, and so will be more compatible with individuals who have reduced cognitive abilities, as compared with synthesised speech which incurs an increased cognitive overhead. In addition to purely cognitive reasons why digitised speech should be utilised, there are also more practical implications.

Remediation Possibilities

In terms of therapeutic impact it is crucial that *spoken* output from ACD's sounds as naturalistic as possible. With digitised speech this level of fidelity is possible. The obvious spin-off of using a natural sounding speech aid is that it will also act as a constantly available speech therapist, which doesn't become tired of constant repetition. With such repetition being associated with the majority of those with language and combined learning difficulties.

Remediation possibilities are enhanced as a result of using digitised speech which is likely to leave some degree of residual processing capacity within the user that can be subconsciously devoted to language learning and enhancement.

Listening to and if applicable repeating natural spoken English is essential for the development of appropriate phonic, or decoding, skills in those with language impairments (Snider 1992, Groff 1990, Tunmer 1990). As speech therapy resources are effectively *rationed* such interactional opportunity, whether person-to-person, or machine-to-person is crucial. Such interactions are also likely to aid other equally important component skills such as, auditory discrimination, short-term auditory memory and sound blending (Westwood 1993). Indeed auditory discrimination, auditory analysis (segmentation), and phoneme blending are now regarded as parts of a more general metalinguistic ability termed "Phonological Awareness" which must be acquired by an individual for meaningful language to develop (Goswami and Bryant 1990, Sawyer and Fox 1991). Many therapists in respect of this prerequisite concentrate on the association of real pictures to their spoken name in a similar way to a "look-and-say" methodology. This methodology is strengthened by findings that young unimpaired language users often learn to name things by listening to the words others use and then reproducing these labels (Ammon and Ammon 1971, Leonard, Chapman, Rowan and Weiss 1983).

Motivational Implications

It is not unreasonable to assume that for language remediation to be effective motivation and attention must also be strong (Westwood 1993 and Walberg 1988). It is proposed that the use of digitised speech increases motivation and that the use of a familiar voice, e.g. that of a parent or carer, increases motivation still further. The speech is seen as familiar and reassuring, in addition to being potentially easier to decode due to its familiarity. Although these assertions have not been experimentally tested, analysis of video taped sessions (Wierman 1992) with Easy Speaker suggest this to be the case. With severely impaired individuals correctly identifying the *speakers* voice, requesting to hear them again, and in some instances requesting to record their own voice/vocal ability or the voice of someone else they know. For example Easy Speaker allows for a scanned picture of a friend to be associated with a piece of that friends speech or whatever. Such pairings can act as *rewards* for successful interactions. For example a scan of the families pet dog can be associated with a dog barking. For some users currently making use of Easy Speaker, pieces of pop music have been recorded and associated with the singers photo or graphic representation. Obviously recording of music with synthesised speech systems is not possible and therefore the potential for increased motivation lost. Use of such *recordings* has had spectacular effects in terms of motivation and on-task attention not predicted by the individuals history or ability. In some cases on-task behaviour has increased to over an hour, with sessions being ceased only on request of the author or other maintainer. As a result the opportunity for language remediation is greatly improved, not to mention potential improvements in attention spans and all round well-being.

Effects on interaction

Use of digitised speech allows for more accurate *spoken* output to be used, that is in terms of age, gender and nationality matching. Many speech synthesisers assume the *voice* of a white middle-aged, middle-class male American. Although many aids still use this formula, logic dictates that remediative effects are not likely to be as good with correctly matched high fidelity speech. Regional accents, age differences, gender differences and cultural inflections can make-or-break interactions, both of the user with the machine and the user with unimpaired and impaired individuals. Any unusualness in the speakers voice can be enough to discourage, stall or even break off interaction (Newman 1982, Scherer 1979 and Alm, Arnott and Newell 1992). Such effects are large enough to be measured when listening to strong regional accents, the implication being that synthesised speech would present a similar barrier (Yule 1986).

Ease of use by parents or care-providers

The use of digitisation means that the person who records the speech or other sounds can operate any such system in a similar way to a tape recorder. That is, without the need to enter into complex programming scripts of text or phonetic transcriptions. Virtually all can operate a the standard controls of a tape recorder, therefore all can record speech etc... with systems such as Easy Speaker. Indeed the ease and immediacy of system preparation can make-or-break the attitude of a parent or care-provider as to its usefulness.

With the advent of care in the community a speech aid that is strongly motivating for the user and easily maintainable by the parents or care-providers is crucial as it is likely to provide a degree of respite care during the periods when the user can operate it independently. Thus benefits on closer inspection are manifold.

SUMMARY

One can argue from theoretical, experimental and practical standpoints why ACD's which produce *spoken* output should where practicably possible utilise digitised speech when in use by an LD population. It must be remembered however that digitised speech cannot be truly generative, as unlike synthesised speech the words or phrases themselves need to be recorded before they can be output. For message based systems for a LD population this is less of a problem.

The hypothesis that synthesised speech by virtue of being an approximation is more cognitively demanding on decoding than digitised speech was supported when subjects with unimpaired language and normal cognitive abilities were evaluated. This must be tempered by the fact that this is in terms of pure cognitive load per se, as digitised speech instructions were complete phrases. That is, Easy Speaker itself for reasons of non-redundancy makes use of concatenated digitised speech on occasion to emulate the generative nature of synthesis systems. It could be argued that this could introduce a slight overhead, over continuous digitised speech.

However the implication is that synthesised speech will incur a greater cognitive overhead for those with decreased language and other cognitive abilities. Confirming further the need to utilise digitised speech were the arguments presented regarding remediation through constant therapy. With digitised speech being likely to have the strongest remediative effects.

A model of the cognition was proposed to account for inherent lower cognitive functioning of LD individuals in relation to language and other related cognitive areas. A specific model of speech processing was proposed to account for the interaction between a slowed cognitive system and overhead imposed by synthesised speech.

Finally, motivation for users and ease of use by parents or care-givers were again supportive of the use of digitisation. Even the argument that machine requirements are too excessive for the use of *mass* digitisation are proving to be somewhat overstated with hardware increasing in power despite decreasing costs with each passing month. The benefits of modern technology should be passed onto the handicapped who can perhaps benefit most, rather than using yesterdays merely "adequate" technology.

In justifying the use of digitised speech with real LD users, off-device improvements in language and cognition would be suggestive of the real world impact its use may have. A study of long term Easy Speaker usage and the impact on off-device language and cognition is discussed in chapter ten.

CHAPTER 8: The development of bespoke evaluation materials

Although off-the-shelf questionnaires could be used for assessing maintainers attitudes and opinions toward Easy Speaker itself, the measurement of users competence and off-device progress required construction of bespoke materials.

In order to measure users competence, or performance, with Easy Speaker a short set of tasks known as the Easy Speaker Assessment Tasks (ESAT) were constructed. These required users to complete set tasks of increasing complexity by making use of a standardised Easy Speaker hierarchy. Results from which could be compared with earlier sessions, the sample norm, and the performance of an unimpaired individual. Such a set of tasks would indicate whether users' were making productive use of Easy Speaker, and whether they were improving with practice.

In addition to pure performance measures it was also necessary to measure off-device improvements in language and cognition. Although set clinical tests do exist which measure language and cognitive improvements, a more behavioural, or adaptive, inventory was required. Two such inventories were considered. Namely the Adaptive Behaviour Scales (ABS) and the Functional Performance Record (FPR). However, both inventories contained areas that were irrelevant, so through an amalgam of the relevant areas from the two a third test called the Easy Speaker Progress Monitoring (ESPM) questionnaire was constructed. This required caregivers to complete one half of the ESPM prior to intervention, and the second some six months into intervention. The second half could then be compared with the first to quantify improvements in language and cognition.

The construction and use of these two hybrid tests were seen as crucial to the measurement of Easy Speaker success. That is, in terms of user competence and off-device improvements.

INTRODUCTION

Unfortunately many previous commercial and research based ACD's have failed to measure users' on device performance in any depth. This has meant that measurements of performance gains have been typically limited to summary statistics such as Words Per Minute (WPM) scores. However, with a highly accurate automated tracking system such as the Automated Response Tracking System (ARTS) built into Easy Speaker this need no longer be the case. As one might expect this not only provides statistics for WPM scores, but also transcriptions of what was output, highly accurate timings, and measures of cursor efficiency etc... In short everything the user does during their interaction with Easy Speaker is monitored and recorded for later analysis. In addition to analyses of what the user does of their own free will this raises the possibility of constructing a battery of set of tasks which can be administered periodically to measure performance improvements on those tasks. This is the goal for the Easy Speaker Assessment Tasks (ESAT), which through increasingly complex set tasks hope to measure user performance.

Performance criterion however are not the only measures that must be taken to reliably assess the impact of Easy Speaker intervention. Other behavioural, or adaptive, measures which are concerned with how the individual interacts with the environment and those around them must also be taken. Such measures can be used more *reliably* than traditional IQ type tests to assess abilities such as language, cognition, motivation etc... They are measures that assess the abilities of the individual in the real world. Although several widely used adaptive behaviour inventories exist, an amalgam of two of the most popular was constructed, with irrelevant items from each being dropped. The hybrid questionnaire, called the Easy Speaker Progress Monitoring (ESPM) questionnaire could be administered prior to intervention to establish a base line and then again after prolonged use. The goal for ESPM is to measure relevant off-device improvements in language and cognition brought about by Easy Speaker usage.

Having outlined the need to assess both on-device performance and off-device improvements in language and cognition each instrument will be discussed in more depth as each will play a major role in assessing the success of Easy Speaker.

MEASURES OF PERFORMANCE: ESAT

Background

The ESAT tasks were constructed with the goal of measuring raw performance and efficiency gains shown by users of Easy Speaker over a relatively short period of time. By re-administering set tasks periodically to users any improvements could be plotted. From these one could infer not only that users could use Easy Speaker per se, but also that they were displaying learning, and formulating and applying high level planning strategies. In addition one could infer that motor skills were also being positively affected.

In choosing the set tasks the abilities of the potential users must be borne in mind. In order to avoid floor and ceiling effects tasks were graded in levels of increasing complexity. With early tasks being very simple to complete as compared with later ones. Tasks were constructed so as to test a range of abilities. Namely the ability to navigate, or traverse the hierarchy, the ability to combine more than one symbol to construct a longer phrase, and at the most complex, tasks where both types of skill were called upon.

Bearing in mind the likely attentional deficits of potential users it was decided to make the ESAT tasks as short as reasonably possible so as to ensure that users always gave their optimum performance. With this in mind a maximum of ten tasks was considered a limit, as these would require a maximum of ten minutes concentration on the part of the user.

Finally, to standardise tasks, the hierarchy with which users would complete the tasks was fixed as the demonstration hierarchy. This remained unmodified until a series of three ESAT task sets had been completed by each user.

The choice of tasks

Eight utterances of increasing complexity were chosen for the user to generate using the demonstration hierarchy. Basic trialling with real users had suggested that about eight sequential tasks provided the best balance between optimum performance and loss of focused attention.

All tasks were utterances which real users might generate by making use of all the features Easy Speaker had to offer. Namely navigation of a number of levels in the hierarchy, selection of symbols, and a combination of the two.

In addition to choosing relevant utterances to generate, a simple scheme for assigning a numerical value for complexity was also constructed. This basically followed the formula shown below :-

$$\text{Complexity} = \begin{array}{c}\text{Number of}\\ \text{levels to}\\ \text{navigate}\end{array} + \begin{array}{c}\text{Number of speech}\\ \text{symbols to}\\ \text{select}\end{array} + \begin{array}{c}\text{Number of different levels}\\ \text{speech symbols selected}\\ \text{from}\end{array}$$

By formally assigning a numerical value for the complexity of each task this enabled tasks of increasing complexity to be constructed. The actual tasks chosen and their complexity ratings are shown below :-

Fig 8.1, Assessment task content and complexity

Task No.	Task category	No. of levels to navigate (a)	No. of speech symbols to select (b)	No. of different levels speech symb. selected from (c)	Utterance to generate	Rating for task complexity (a+b+c)
1	Navigational	1	1	1	Meals	3
2	Navigational	2	1	1	Summer	4
3	Navigational	3	1	1	Trainers	5
4	Navigational	4	1	1	Muscles	6
5	Constructional	2	2	1	Ham - salad	5
6	Constructional	3	3	1	Read - library - book	7
7	Combination	3	2	2	Is it - raining	7
8	Combination	3	3	3	Get the nurse - sore - back	9

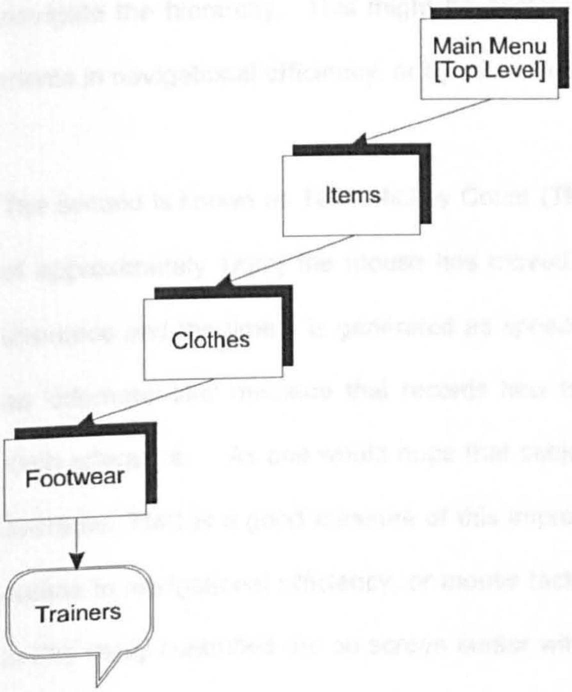
The goal of the navigational tasks (1-4) was to assess whether users could navigate to, select and output single symbol utterances. In general they would be required to generate these single symbol utterances in response to a simple question, or statement posed by the administrator. For example, "What sort of shoes are those you're wearing ?". The exact nature of these instructions was varied to take account of the abilities of a particular user.

The constructional tasks (5-6) were designed to assess whether users could generate logical multi-symbol utterances. For example, they might be asked which sort of salad they liked, gently steered toward "Ham salad" and asked to say "Ham salad". That is, reproduce it using Easy Speaker.

Finally, the combination tasks (7-8) were the most complex utterances users would be likely to generate if they were using Easy Speaker as their primary means of communication. That is, multi-symbol utterances generated by selecting appropriate symbols from various levels within the hierarchy. Again with appropriate prompting they were requested to generate these ESAT tasks. Successful completion of these tasks would be likely to suggest that if those same users were completely non-vocal they would be able to make use of Easy Speaker as their primary means of communication. The speed and efficiency they might do so is crucial and so these are the precise measures ARTS takes by recording constructional latency and amount of cursor movement.

As an aid to the maintainer, or administrator, who would oversee completion of the ESAT tasks they were represented diagramatically to avoid confusion.

Fig 8.2, An assessment task represented to the administrator as a tree diagram



In this case the subject was required to navigate to, select, and finally generate as digitised speech the word "Trainers".

Key

= A level in the hierarchy

= A speech symbol at a given level in the hierarchy

Choice of ARTS performance criterion

Although Easy Speakers in-built tracking system records virtually all aspects of a users interaction, for the purposes of measuring performance gains it was found that three key measures are sufficient. Two of the measures are recorded verbatim, with the third being calculated by ARTS.

The first of the two measures is known as Total Time (TOT T) and is a measure, in seconds, of the elapsed time between starting to construct an utterance and the time when it is generated as speech. A decreasing TOT T would suggest an increase in the speed with which the user could navigate the hierarchy. This might be explained purely as a speed improvement, by improvements in navigational efficiency, or by a combination of the two.

The second is known as Total Mickey Count (TMC) and is a measure of how far in mickeys (units of approximately 1mm) the mouse has moved on the desktop between starting to construct the utterance and the time it is generated as speech. The mickey count can simply be thought of as an 'odometer-like' measure that records how far the mouse cursor has travelled in constructing each utterance. As one would hope that subjects develop more efficient navigational strategies overtime, TMC is a good measure of this improvement. A decreasing TMC would suggest an increase in navigational efficiency, or mouse tack. It is worth pointing out the although all subjects in this study controlled the on-screen cursor with the cursor keys, using a measure such as mickeys is still valid as the keys directly emulated the mouse at a hardware level.

The final calculated measure is an estimate of how many words per minute the subject would be likely to construct and generate based on their TOT T for the last utterance and the number of words it contained. Their words per minute estimate varies with each utterance they construct and is calculated using the following formula :-

WPM = 60 sec's / (TOT T / number of words in utterance)

* N.B. - TOT T was measured in seconds and hundredths

For example, to manually calculate the WPM score for "get the nurse sore back" the following calculation would need to be performed :-

60 / [137.58 / 5] = 2.18 WPM

The estimated WPM simply provides a 'rough and ready' measure of users ability and the potential of Easy Speaker itself. One must bear in mind that as any speech symbol could represent several sentences of speech and a relatively high wpm might be misleading when viewed in this context.

Administration

The administration of both the ESAT tasks and ESPM questionnaire will be described as they might in a standard test manual. That is, the exact procedure needed to be followed in order to administer them is outlined so that a third party may apply them as necessary. All procedures outlined were those followed by the author when applying them.

Before a users progress can be assessed using ESAT tasks it is envisaged that they should have clocked up around eight hours of Easy Speaker interaction. This will enable them to become accustomed to the controls, method of operation, and the layout of the standard hierarchy.

In order to measure performance gains accurately users should not intentionally, or unintentionally, over practice any of the eight ESAT tasks or their components. Thus it is more likely that performance will reflect actual performance gains in Easy Speaker usage rather than a fine honing of skill with the test materials per se. In order to ensure this, tasks should be completed once each week over a consecutive three week period.

Within each of the three sessions users should be allowed the first 20 minutes or so to re-orientate themselves with Easy Speaker and its controls. This should take the form of free/semi-structured interaction before being asked to construct and generate the eight ESAT task utterances shown in Fig 8.1 on page 190.

After the user had 20 minutes or so of interaction with Easy Speaker to reorientate themselves, they should be asked to settle themselves and to construct each utterance in turn as quickly and as efficiently as they their ability allows. The user should simply be told the target utterance they are required to construct and generate. They should not constantly be prompted as to the best way to achieve this in terms of navigational strategies or cursor tack etc... That is, one should not depart from the prompting scheme outlined below. Between each task the user should be told to settle themselves before being asked to construct and generate the next task. Immediately prior each task the system should be reset so the user is located at the top level in the hierarchy with the cursor located in the centre of the screen. This resets ARTS and therefore acts as a datum from where the tracking system begins tracking the users timings and cursor movements etc...

As users would be *blind* in the first session to the ESAT tasks the administrator should demonstrate how to complete each task before the user attempts that task. Users should be given varying levels of support dependent upon which session they are being evaluated under. Support levels are based upon the following scheme :-

Fig 8.3, Levels of support which should be given to users during completion of ESAT tasks

Week No.	Level of support given to users during ESAT tasks
1	***Full support*** Users should be given a demonstration of how to complete each ESAT task in turn. After each demonstration they should be asked to navigate the hierarchy in order to complete that task. Full prompting on which link or speech icon to select next should be given to users.
2	***Medium support*** Users should be prompted on which link or speech icon to select next when they become either stuck or confused.
3	***Minimum support*** Users should only be prompted as a last resort on which link or speech icon to select next when they become either stuck or confused, and cannot continue.

No users should be allowed to totally fail any ESAT task regardless of how long it takes or how many attempts are needed. This stance is taken so as to avoid any negative implications resulting from task failure. During piloting all ESAT tasks proved to be so well matched to users' abilities that no user actually required more than the most basic of prompts in order to complete a given task.

Prior to administration it is envisaged that the administrator should have ensured that all users should have completed tasks of equivalent task complexity during their familiarisation period.

Once a user has completed all ESAT tasks for that session, the session should be carried on to a conclusion as normal concentrating on semi-structured and structured interaction. Only after the session is fully complete should analysis of ARTS data begin.

Analysis of ARTS data for the ESAT tasks

In-depth analysis of the data generated by the user during completion of ESAT tasks is basically a matter of preference for the administrator. This is due to the fact that there are no norms as such for performance to be evaluated against. Performance is simply evaluated within users. The only norms that can be applied are norms for the group of users currently undergoing assessment, and the best case norms. However general guidelines for assessment are provided below.

As ARTS data is stored in a Comma Separated Value (CSV) file this means that it can be read directly into any spreadsheet, e.g. MS Excel, Lotus 123 etc... and analysed accordingly. Suggested summary analysis should be to compare :-

TOT T	Total Time
TMC	Total Mickey Count
WPM	Words Per Minute scores

Performance differences between sessions can best been seen by calculating means and then plotting them to give a graphical representation. For example, when means and Standard Deviations are calculated for TOT a table similar to the one below should be constructed.

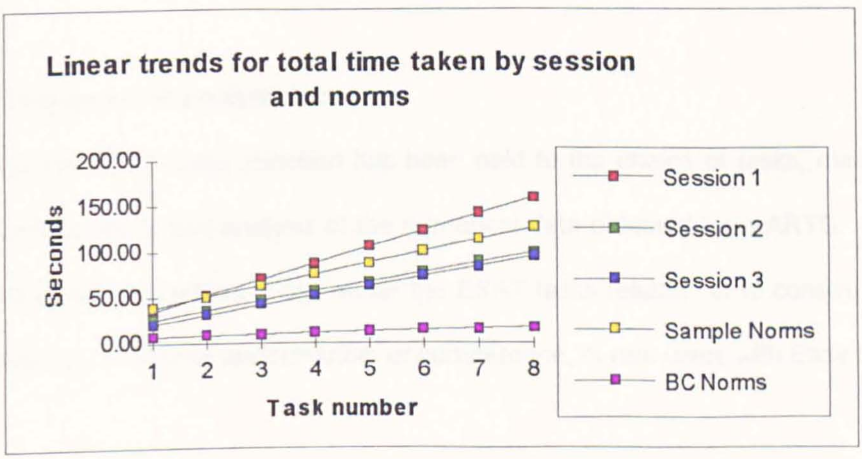
Fig 8.4, An example of how to calculate summary statistics for the ESAT tasks

Task No.	Transcription of utterence subject had to construct	Session 1 TOT T	Session 1 TMC	Session 1 WPM	Session 2 TOT T	Session 2 TMC	Session 2 WPM	Session 3 TOT T	Session 3 TMC	Session 3 WPM
1	meals	54.49	138.00	1.10	28.84	101.00	2.08	25.81	78.00	2.32
2	summer	87.60	118.00	0.68	36.53	101.00	1.64	41.46	83.00	1.45
3	trainers	73.99	149.00	0.81	57.06	106.00	1.05	47.01	94.00	1.28
4	muscles	116.72	165.00	0.51	112.21	138.00	0.53	91.34	130.00	0.66
5	ham/salad	104.09	204.00	1.15	61.85	143.00	1.94	67.61	131.00	1.77
6	read/library/book	144.18	198.00	1.25	86.73	158.00	2.08	73.66	136.00	2.44
7	is it/raining	117.82	236.00	1.53	83.98	139.00	2.14	88.04	147.00	2.04
8	get the nurse/sore/back	216.74	234.00	1.38	115.95	183.00	2.59	111.77	190.00	2.68
	Mean	114.45	180.25	1.05	72.89	133.63	1.76	68.34	123.63	1.83
	SD	49.89	44.31	0.35	32.40	29.42	0.66	28.83	37.44	0.68

As can be seen the user above became faster (TOT T) with each session, his efficiency improved (TMC), and he progressively generated a higher number of words each minute (WPM).

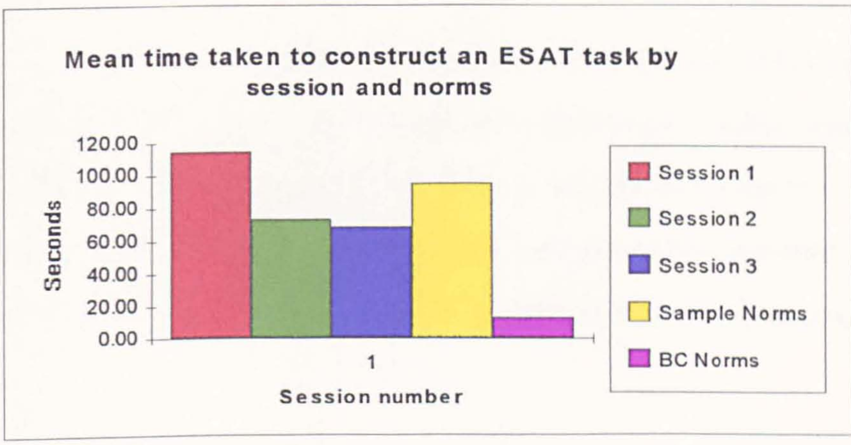
As the raw data may be difficult to visualise, the author plotted linear trends as opposed to raw data.

Fig 8.5, Using linear trends as opposed to raw data in summary performance plots



By plotting trends any improvements in performance usually become clearly visible. However, if they do not, or if further analysis is required these improvements can be seen by plotting session means using a simple bar chart.

Fig 8.6, Using a simple bar chart to plot performance improvements over sessions



To contrast a users performance relative to others of similar ability, sample norms could be constructed. Alternatively best case (BC) could be constructed for an unimpaired individual when completing the same tasks.

Relevant statistical analysis of the administrators choice can then be applied to assess whether improvements between sessions were significant. A users performance could also be compared with the sample and best case norms for the same tasks.

Summary of measures of performance

In constructing the ESAT tasks attention has been paid to the choice of tasks, measures of performance, administration, and analysis of the numerical data obtained from ARTS. Hopefully this will enable the author and others to use either the ESAT tasks reliably, or to construct ones with a similar methodology to assess performance, or competence, of real users with Easy Speaker.

MEASURES OF OFF-DEVICE IMPROVEMENTS IN LANGUAGE AND COGNITION: ESPM

Background

The ESPM questionnaire was constructed with the goal of measuring off-device improvements primarily in language and cognition over a prolonged period of Easy Speaker usage. By administering the ESPM questionnaire to two of each users immediate caregivers prior to intervention, and then again after six months any off-device improvements could be assessed. In addition, by administering the ESPM to two caregivers inter-rater agreement on the users progress could also be assessed.

As opposed to more formal assessment instruments a more informal adaptive behavioural inventory was sought. That is, one that focused on what a user could do, as opposed to what they couldn't, e.g. an adaptive behaviour scale versus a standard IQ test. Although numerous such inventories do exist any in isolation would contain many irrelevant items. For example, the Adaptive Behaviour Scale includes many items on personal care and hygiene, in addition to those on language and cognition. With this potential drawback in mind it was decided to produce a hybrid questionnaire by combining relevant items from two of the most popular adaptive behaviour scales. Namely the Adaptive Behaviour Scale and the Functional Performance Record.

However, unlike standard questionnaires that are re-administered twice, each half of the hybrid ESPM questionnaire was not identical. This was due to the fact that a positive or negative change over baseline was the goal for measurement, and not a difference in rating per se. This meant that smaller changes were likely to be registered on the second questionnaire than otherwise might be the case.

The choice of questionnaire items

Both the ABS (American Association on Mental Deficiency) and FPR-16 (NFER-Nelson) contained around 200 items apiece, with some being more relevant to assessing Easy Speaker intervention than others. Both contained not only items on language and cognition, but also items concerned with other adaptive behaviours such as feeding and toileting etc...

With relevance to constructing the ESPM questionnaire it was decided to formulate domains that were clearly relevant in the measurement of off-device improvements. Six key adaptive behaviour domains were considered crucial. Namely :-

❶ Expressive language	❹ Initiative
❷ Receptive language	❺ Attention
❸ Social language development	❻ Memory

By selecting items from the ABS and FPR that fitted easily into these domains a total of 69 relevant items were selected.

However, both the ABS and FPR make use of differing response scales. The ABS uses a simple check box methodology, whereby each item has a list of statements below it that are checked as being applicable in relation to the item in question. Whereas the, FPR makes use of a Likert scale that can be variable with each item. Obviously creating a hybrid is not a simple matter of combining items without modification.

For example from the FPR, section 210 deals with "Speech and Language Production". The first item(s) is presented as :-

Fig 8.7, Item 1 from section 210, Speech and Language Production, of the FPR

What problems of speech production did s/he have during the last week ?

Codes

slight problem = 3	not applicable = 9
marked problem = 5	unkown = 0

lisp ☐
stammer ☐
speaking too rapidly ☐
voice too faint/soft ☐
over loud speech ☐
word finding difficulty (dysphasia) ☐
slurring (dysarthria) ☐
other articulatory difficulties ☐

In order to combine the two set of items, and retain their face validity, what was needed was a third scale which could be used with both. The most relevant scale was decided to be a seven point semantic differential scale ranging, never - sometimes - always. In order to fit the 69 items to this scale some slight rewording was required.

For example all items in the above FPR example, were assigned to the ESPM domain of expressive language. Item six for the pre-intervention section of the ESPM became :-

Fig 8.8, Item 6 from the pre-intervention section of the ESPM questionnaire

Does the client have word finding difficulty (dysphasia) ?

☐	☐	☐	☐	☑	☐	☐
Never	Rarely	Occasionally	Sometimes	Often	Frequently	Always

In the example above the item has been reworded to read "Does the client have word finding difficulty (dysphasia) ?", rather than the FPR original, "word finding difficulty (dysphasia) ?".

Despite the change in scale obviously both items still assess the same criterion within the same domain. In addition as data from the hybrid ESPM would not be compared with standardised norms for either the ABS or the FPR, then changes in response scale were relatively unimportant. What was is the content of the items themselves.

Having constructed a 69 item questionnaire split into six domains, the second half of the questionnaire needed to be constructed. However, as mentioned briefly this was not simply a matter of using exactly the same items with the same scale. The was due to the fact that the first half of the ESPM was used to establish a baseline and the second half was used to measure changes.

As such it was decided that caregivers would have their baseline response restated in relation to each item, and then be asked whether they though the user had changed in ability. Again a seven point semantic differential scale was used. This ranged from, much worse - no change - much better. Using the example in Fig 8.8, its equivalent item from the second half of the ESPM, or post-intervention, is shown below :-

Fig 8.9, Item 6 from the post-intervention section of the ESPM questionnaire

When asked previously "Does the client have word finding difficulties (dysphasia) ?" you stated ....*Often*......... Are they now :-

☐	☐	☐	☐	☐	☑	☐
Much worse	Worse	Slightly worse	No change	Slightly better	Better	Much better

By using a standard semantic differential scale this meant that all positive responses were to the right of the scale enabling more straight forward analysis. In the example above, as the response post-intervention is to the right of the scale this would suggest a positive change for that user for that item. In this case, that their word finding difficulties, or dysphasia, had decreased.

Administration

Administration of the ESPM questionnaire is relatively straight forward. The first half of the ESPM should be given to at least two of each potential users immediate caregivers prior to the start of Easy Speaker intervention. This is done to establish a baseline against which any improvements can be compared. Once complete this should be stored for safe keeping.

After a reasonable period, ideally not less than three months after the start of Easy Speaker intervention, the second half of the ESPM should be prepared to be given to the same caregivers as rated that user initially. Preparation involves writing in the caregivers initial responses in the gaps provided for each item. This is done so that the caregiver can gauge any improvements against their original baseline ratings.

Caregivers who complete the ESPM should not, where possible, be involved in the actual process of Easy Speaker intervention. That is, they should not be a maintainer, who may unwittingly have a vested interest, and unduly affect responses. The goal is to gain as objective a rating, both pre and post intervention as possible.

Instructions for caregivers are provided on the cover of both halves of the ESPM. Examples of correctly completed items are also given. The actual cover instructions from the pre and post intervention EPSM questionnaires are shown below.

Fig 8.10, The pre-intervention EPSM cover and caregiver instructions for correct completion

Easy Speaker Questionnaire

Please place a tick in the box which best describes the clients current level of ability, e.g. :-

Does the client ever wear red coloured clothing ?

☐	☐	☐	☑	☐	☐	☐
Never	Rarely	Occasionally	Sometimes	Often	Frequently	Always

If you can't comment on the ability described simply leave all boxes blank.

Client Name (ID): ....

Date: ....

Location: ....

Observed by: ....

Position or relationship to client ....

How long known to client ....

Fig 8.11, The post-intervention EPSM cover and caregiver instructions for correct completion

At an earlier date you were asked questions which described the clients level of ability. In this questionnaire you will be reminded of you answers you gave and asked to comment on whether the client has worsened, stayed the same, or improved etc...

For example :-

When asked previously "Does the client ever wear red coloured clothing ?" you stated that the client ..... *Sometimes* ..... wore red clothing.

Do they now wear red clothing :-

☐	☐	☐	☐	☐	☑	☐
Very much less	Much less	Slightly less	The same	Slightly more	Much more	Very much more

Please place a tick in the box which best describes the clients level of improvement. If you can't comment on the ability described simply leave all boxes blank.

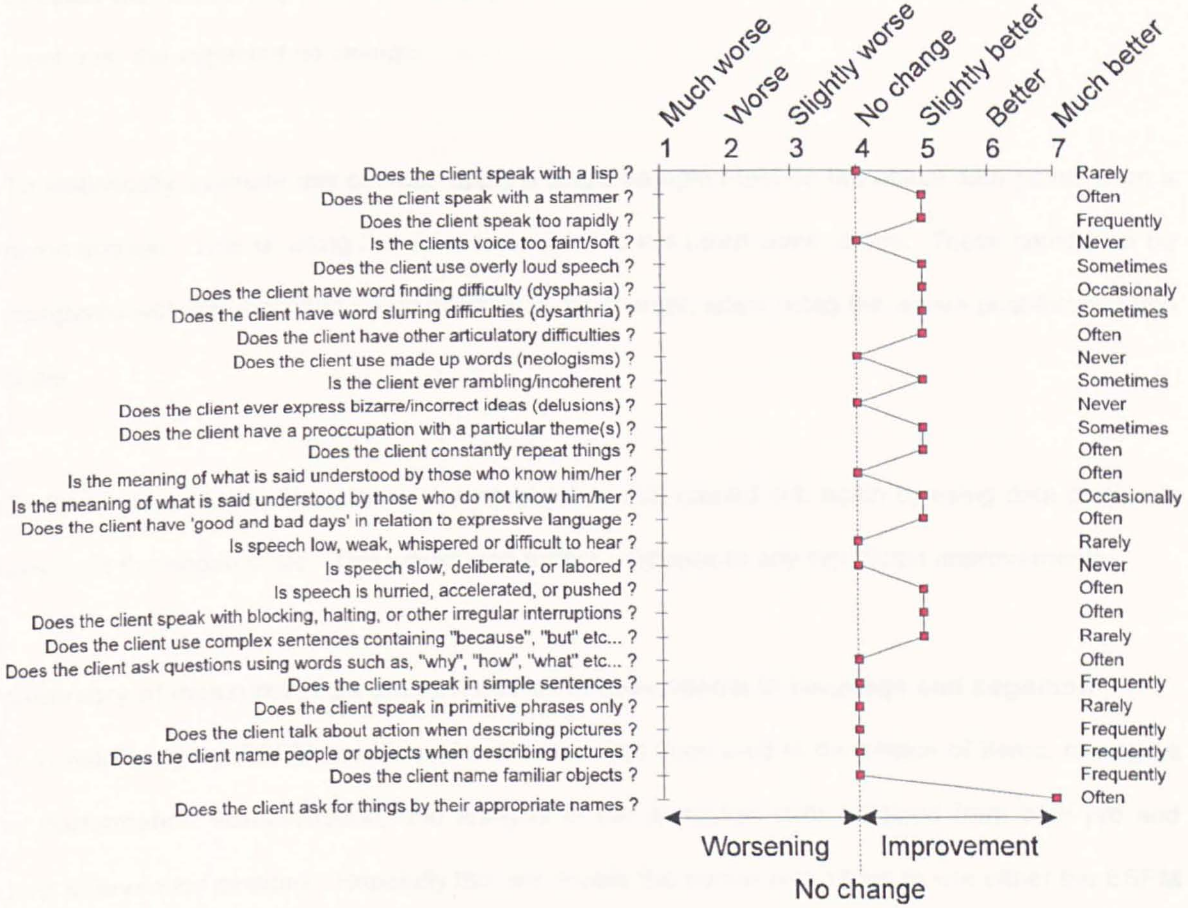
When both halves of the ESPM questionnaire are complete they are ready for analysis.

Analysis of ESPM responses

Analysis of EPSM responses, in common with the questionnaire design itself, is not simply a matter of comparing test-retest responses. Obviously this cannot be done as each half of the questionnaire uses a different semantic differential scale. However, analysis is relatively straight forward.

Before carrying out any form of statistical analysis it is useful to graph the actual response data for each user by means of a modified line graph.

Fig 8.12, ESPM ratings for Expressive Language represented graphically



By use of such graphs it is simple to see any improvement as peaks will be to the right of the central axis, which indicates no change. If the user worsens these peaks fall to the left. As can be seen from the graph all comparisons are contrasted with no change.

Actual questionnaire items are displayed on the vertical axis, with pre-intervention responses shown on their immediate right. This indicates which responses constitute a baseline for that particular user. The top axis represents the semantic differential scale for the post-intervention half of the ESPM, along with a numerical, Likert type, value.

By applying such a visualisation technique to the caregivers responses any improvements can easily be seen. In order to check whether any suggested improvements are significant one can compare the expected outcome of Easy Speaker intervention with that obtained within each domain. That is, if Easy Speaker was exerting little influence over the abilities of users one would expect consistent ratings of no change in EPSM items at post intervention. However, if Easy Speaker did prompt improvements in a given domain then there would be a significant improvement over the expected no change.

To statistically evaluate this one can apply a single sample t-test on the actual data points from a given domain. That is, using the values assigned to the Likert scale above. These could then be compared with the expected value which is 4, no change, when using the above post-intervention scale.

Further analysis regarding inter-rater agreement can be carried out, again by using data point values from the above scale. This would lend further credence to any significant improvements.

Summary of measures regarding off-device improvements in language and cognition

In constructing the ESPM questionnaire attention has been paid to the choice of items, measures of performance, administration, and analysis of the numerical data obtained from both pre and post intervention sections. Hopefully this will enable the author and others to use either the ESPM questionnaire, or to construct ones with a similar methodology to assess off device improvements in language and cognition of real users of Easy Speaker.

SUMMARY

The rationale behind the two bespoke evaluation materials used to assess the success of Easy Speaker intervention with users has been outlined.

The first, the ESAT tasks are used to assess users raw performance when using Easy Speaker, and are used to plot improvements in speed and efficiency. From these one can infer that learning has taken place, high level planning strategies are being employed, and that motor co-ordination has been honed.

The second, the ESPM questionnaire is completed by caregivers both prior to and post Easy Speaker intervention to measure off-device improvements in language and cognition. From caregivers responses one can calculate whether significant improvements in any of the test domains have occurred as a result of Easy Speaker intervention. In addition to lend credence to any findings levels of inter-rater agreement can also be calculated.

The ESAT tasks will be used in chapter 9, to assess real users' competence with Easy Speaker. That is, whether they can actually learn to use the system productively, and to plot levels of improvement. The ESPM questionnaire will be use in chapter 10 to assess the impact of long term Easy Speaker usage on language and cognition in real users.

CHAPTER 9: Easy Speaker in use: An assessment of user competence

Although Easy Speaker's development, implementation and rationale had been viewed favourably by both potential users and professionals alike, it was not demonstrably known whether users could acquire the skills necessary to operate and potentially benefit from the system. In order to examine these issues a three week sub-study of four moderately Learning Disabled, language impaired users was conducted. The study was specifically designed to assess patterns of skill acquisition and competence using the Easy Speaker Assessment Tasks (ESAT) discussed in chapter eight. Assessment focused on :-

1. Whether users could successfully construct set task utterances which required them to navigate a large and relatively complex hierarchy.
2. Whether users developed more efficient navigational strategies over time.
3. Whether construction speeds improved.
4. Whether users levels of competence improved so significantly as to approach or surpass the average competence shown by the sample as a whole (sample norms).
5. Whether users levels of competence improved so significantly as to approach or surpass the average competence shown by an unimpaired user (best case norms).
6. Whether users could construct meaningful multi-symbol utterances.

Successful completion of ESAT tasks, and significant performance gains over sessions would suggest that potential users can acquire the necessary skills required to operate an apparently complex hypermedia based Augmentative Communication Device. Crucially it would also suggest that the hypermedia metaphor and that the physical aspects of the interface design were appropriate. If users could construct multi-symbol utterances, of their own volition, this would suggest that Easy Speaker could be used as a primary means of communication as originally intended.

INTRODUCTION

Although use of any interventional strategy which hopes to gradually improve a users language functionality and general cognitive abilities can *only* be measured over the long-term, a short three week sub-study was carried out by the author. The aim of this study was to assess whether users could learn to *use* Easy Speaker successfully. That is, to assess whether they could acquire the skills needed to successfully navigate a large and relatively complex hierarchy, to examine navigational strategies used, and to quantify any performance gains. Simply, the ability to drive the device, and produce meaningful output.

Despite an earlier single-subject study using Easy Speaker for DOS proving to be a major success, grading of this success could be criticised for its subjectiveness (Rostron, Plant and Hermann 1994). In this particular case formal performance assessment could not be carried out due to the severity of the subjects learning disabilities, level of communicative ability and unknown level of visual impairment. The study did however suggest the hypermedia metaphor employed by Easy Speaker was appropriate for potential user populations identified and that the psychological rationale behind the approach was sound (see chapter 4 for a full discussion). It also suggested that the on-screen representation of the underlying hypermedia metaphor was appropriate from a HCI prospective as the subject rapidly learned how links between levels were represented despite the difficulty of conveying this to him explicitly. What was needed however was a formal way of measuring improvements in skill acquisition.

In quantifying raw performance gains Panyan (1984) points out that many studies in the area did and indeed continue to "... lack appropriate experimental design ... which in turn makes it difficult to draw any firm conclusions ... as to their success". Such an assertion is hardly surprising considering that *no* commercial devices offer any form of user response tracking. Consequently, objective assessment of on-device performance is near impossible in everyday settings. Assessment is in the main left to subjective judgement accompanied by the use of a stopwatch.

With the Automated Response Tracking System (ARTS) built into Easy Speaker, monitoring of skill acquisition and competence is made both objective and substantially easier to implement. Indeed ARTS enables Easy Speaker to be used as an assessment tool in its own right, as each utterance generated is recorded verbatim along with relevant performance measures. Following Monk's (1984) suggestion that, "The most important measures of overall efficiency are time to complete the task and some quantification of correctness.", measures of skill acquisition in this study focused on how long a user took to construct a set utterance, and how far the cursor was moved during that construction. Both of which are measured accurately by ARTS as utterances are constructed (see chapter five for a full discussion of ARTS).

Using such recorded measures allows for any improvements to be assessed alongside earlier sessions for that user, for performance curves to be plotted and learning rates to be inferred (Chen and Bernard-Opitz 1993). In addition to using recorded measures, an estimated Words Per Minute score can also be computed. By tracking performance responses for set tasks the resulting data provide an objective way of drawing conclusions as to the suitability of the hypermedia metaphor and the physical attributes of the software interface. One can infer that if users could make productive use of the system and that their ability improved over time the system was well matched to their abilities, or is synergistic. If the hypermedia metaphor proved ill founded, or was simply too complex, then this would be reflected in users skill acquisition rates and poor performance on set tasks. Simply, their performance would not improve significantly over time. The study then is designed to assess the suitability of the Easy Speaker ethos and is not a comparison of interface strategies per se, nor is it intended to be a comparison of the relative merits of commercial devices.

CAN USERS USE EASY SPEAKER

The Paradigm

The paradigm being followed is a modified version of Chen and Bernard-Opitz (1993) procedure used for comparing learning and behavioural rates with regard to personal and computer assisted instruction for children with Autism. Under this modified paradigm users are required to construct a set of eight increasingly complex task utterances (ESAT tasks) once each week over a three week period. The three sessions recorded by ARTS are each analysed in relation to the criterion's outlined below in order to quantify performance gains, to infer learning rates, and ultimately to gauge the suitability of the hypermedia metaphor and interface design for everyday communication :-

1. Time taken to construct each set task utterance (constructional speed or latency)
2. How far the cursor was moved whilst constructing each set task utterance (efficiency)
3. Estimated Words Per Minute based on each set task utterance

From these criterion's three further measures are calculated :-

4. Mean time taken to construct and generate all task utterances
5. Mean distance the cursor was moved in constructing all task utterances
6. Estimated mean Words Per Minute based on the construction of all task utterances

Measures 1-3, provide data on whether the set tasks complexity is loaded correctly, i.e. whether the complex tasks take longer to complete than the simple. Measures 4-6, on the other hand allow comparisons to be made between each of the three sessions. Over the three sessions the hope is that significant performance and efficiency gains will be shown by each user as compared with their initial performance at session 1, or baseline.

Although the study did not attempt to match users in any way other than their general degree of Learning Disability, the use of two sets of calculated norms enables crude comparisons between users to be made. Sample norms gave the average performance for a given criterion over the three sessions for all users. This enabled a statement to be made on whether the user was performing above or below average for the sample. Similarly, best case norms provided the average performance for an unimpaired individual. This enabled a statement to be made on the users performance in relation to that of an unimpaired individual. In addition to purely performance based measures, qualitative assessment of users output during the assessment period as a whole will be considered.

Design

The design followed is based upon a repeated measures single-subject design with parallel replication (Barlow and Hersen 1984). Due to the variability in user ability and impracticability of matching, a repeated measures design was employed. The impracticality of subject matching is evident when comparing simple physical traits such as uncorrected vision. One subject had poor uncorrected vision which obviously affected his performance as compared with his peers who had good corrected vision. Within such a heterogeneous user population many physical and cognitive traits display similar variation, e.g. vision, hearing, motor co-ordination, attention and memory span, test based IQ measures etc... Fortunately, these traits are more-or-less constant within the same individual. Therefore baseline performance measures taken during the first session act as a yardstick to compare any performance gains displayed by the same individual during later sessions. Each user therefore acts are their own control which should in turn lead to increased internal validity.

As the ESAT tasks offered a standard way of assessing users progress the set eight tasks they contained were used over a three week assessment period. The first week established a baseline by which progress could be measured. This was done in preference to the introduction of new tasks each week for which there was no established baseline as such.

External validity is enhanced by utilisation of parallel replication. That is, for findings to be generalisable, comparisons across groups have to be valid, e.g. age, sex, IQ, memory span etc... However, in this case due to the heterogeneous nature of the sample, and the for that matter the population from which they were drawn, each positive parallel replication adds to the external validity collectively.

Users

Users consisted of two male and two female moderately learning disabled adults, with a mean Chronological Age of 28 years 3 months, and a mean Age Equivalency of 5 years, when assessed using the British Picture Vocabulary Scale immediately prior to Easy Speaker intervention. A summary of BPVS scores is shown below :-

Fig 9.1, A summary of users' BPVS scores

User ID	Gender	Chronological Age	Age Equivalent	Confidence Interval	Raw Score	BPVS Form
U1 (PS)	Male	32	5-11*	5-5 to 6-5	55	Long
U2 (AW)	Male	21	3-7	2-11 to 4-3	8	Short
U3 (KC)	Female	34	6-7	6-1 to 7-1	61	Long
U4 (CP)	Female	26	4-0	3-4 to 4-8	9	Short
Mean		28-3*	5-0	4-5 to 5-7		
SD		5-11				

* Shown in years and months

All users in addition to having a Learning Disability, had relatively poor communicative abilities with the exception of communication with immediate caregivers as one might expect. All had deficits in attention, movement and motivation. No user could read or write, nor had any ever been taught any form of augmentative communication system, e.g. Makaton etc... All users were non-independent living with the exception of U4:CP who lived in very closely supervised sheltered accommodation. U4:CP despite her relatively poor BPVS scores was the most functional in terms of adaptive behaviour however. All had previously seen speech and communication therapists over a prolonged period at some stage during their life with varying degrees of success. None were currently receiving speech therapy input, and on average the last input had ceased around 5-10 years prior to Easy Speaker intervention. Both males were currently involved in behaviour management programmes with clinical psychologists due to their severe Challenging Behaviour (CB) and self-abusive behaviours.

Where practical Easy Speaker use attempted to fit in with these on-going programmes as an aid to help individuals communicate their needs and emotions. For other users its use was intended to solely improve communicative competence. During this particular study users made use of Easy Speaker for two hours per week over a seven week period, of which the last three were used for assessment in addition to on-going work. For a full breakdown see fig 9.2 on page 216.

Where further assessment information was available, this is presented along with brief descriptions provided by the users immediate caregiver :-

User 1:PS

Male aged 32 at date of participation

WAIS

	Verbal IQ	52
	Performance IQ	54
	Full Scale	50

British Picture Vocabulary Scale (Long Form)

	Raw Score:	55
	Age Equivalent:	5 years 11 months
	Confidence Interval:	5-5 to 6-5

Carrow Test for Auditory Comprehension of Language

	Raw Score	83
	Age Equivalent	6 years 2 months

U1 had a normal birth but with a history of delayed milestones and is classified as having moderate learning handicap, and showed marked hyperactive behaviour as a child. U1 has been tried on and still receives various neuroleptics in an attempt to control his behaviour. These include Chlorpromazine, Diazepam, and Benperidol. U1 is also prescribed Triclofos (sedatives) to help combat sleeping problems.

U1 also receives anti-Parkinsonian medication to help reduce side-effects of neuroleptic medication. However U1 continues to show fine tremors of both hands. Gross motor skills are intact. Fine movements and co-ordination present a degree of difficulty when he is anxious. He also continues to use only thumb and index finger to prick objects up. Both hearing and vision are in the normal ranges.

U1 has a friendly disposition and appears to like being in the company of staff and other residents, so long as the group remains fairly small that is. He makes little eye-contact and appears anxious in conversation. At times he is quite anxious about everything he is involved in. Displays several stereotyped behaviours which include hand and finger flicking, genital touching, aggressive acts toward others, and self-injury; collectively labelled Challenging Behaviour (CB). U1 is also prone to abscond when not under close supervision.

U1 has an average attention span, and has an above average memory when assessed both over the short and long term. As a result he has the ability to learn new skills relatively quickly; a fact that is not reflected by his IQ and MA comparison scores. To his detriment U1 tends to rush tasks and as a consequence unsuccessfully complete them. This is probably a result of his general tendency toward over-activity/hyperactivity. U1 has good perception of time and possesses adequate monetary skills.

U1 has limited literacy skills. He can write his own name and copy write accurately. He cannot recognise or write words other than his own name. He recognise the letter "P" but no other single letters.

U1 does not initialise conversation. His speech is repetitive and enquiring. Although his speech is not clear he can generally make himself understood to immediate caregivers.

U1 has problems with pronouns and the past tense. Phonetic substitutions are also evident, with "t" and "d" substituted for "k" and "g" etc...

Although technically not dysfluent, U1's rapid speech rate leads to poor intelligibility. U1 tends to use a monotone telegraphic style of speech regardless of situation.

Self-monitoring skills are poor with U1 not recognising when a deterioration in intelligibility is occurring or being aware of the difficulties being experienced by the listener.

User 2:AW

Male aged 21 at date of participation

Stanford-Binet IS (L-M form):

MA:	4 years 9 months

British Picture Vocabulary Scale (short form):

Raw score:	8
Age equivalent:	3 years 7 months
Confidence interval:	2-11 to 4-3

U2 has been learning disabled since birth due to genetic abnormalities, and is classified as having moderate learning handicap. Although U2 is epileptic, fits are scarce and have been well controlled by anti-convulsents. As U2 is also prone to periods of over-activity, which can occasion border on hyper-activity, Triclofos (a mild sedative) is prescribed in an attempt to moderate mood and activity swings.

In addition to mental handicap U2 has a moderate reduction in extents of body movement which is accompanied by persistent jerkiness. Limited extents of movement are most pronounced in U2's wrist movements, which are relatively narrow in range. In an attempt to combat these problems U2 attends regular physiotherapy sessions.

U2's vision is generally poor, but is corrected with glasses. Hearing is in the normal range.

U2 has a limited attention span, of approximately 1..2 minutes in length and is easily distracted by extraneous external events. Indeed this is probably a consequence of U2's proneness to general over-activity. However, U2 can concentrate on the task in hand for short periods.

U2's memory span and retention are somewhat to the lower end of the normal range.

U2 also displays several maladaptive behaviour patterns, which have been partially corrected with the aid of behaviour therapy. Some of which could be considered obsessive in nature, and are again covered by the umbrella term Challenging Behaviour.

Although a fairly competent language user, U2 tends to change conversational topic mid-flow, probably again as a result of proneness to over-activity. U2 uses nearly normally formed sentences, of normal length and can keep up a conversation relatively easily. However, U2's pronunciation and articulation is slightly below par, and combined with rapid rate of speech production conspire to make U2 less intelligible than otherwise might be the case.

User 3:KC

Female aged 34 at date of participation

British Picture Vocabulary Scale (Long Form)

Raw Score:	61
Age Equivalent:	6 years 7 months
Confidence Interval:	6-1 to 7-1

U3 has been learning disabled since birth due to genetic abnormalities, and is classified as having moderate learning handicap.

U3 is currently taking sedatives which can have a pronounced affect on her motor skills and speech on a day-to-day basis. Typically even when sedative free her speech is very slow, laboured, and typically disjointed. Communication usually takes place in a 'stop-start' telegraphic mode, with the effort required to produce the next word clearly visible on her face.

She has under gone several operations on both hands to correct birth deformities in her fingers. These mean that fine hand and finger control is difficult.

Unfortunately her vision remains uncorrected, despite several attempts to remedy the situation. She cannot read or write and makes use of no alternative communication strategies.

Her attention span is in the normal range, although her longer-term memory is prone to deterioration.

She is enthusiastic and enjoys socialising. Due to communication difficulties she has attended no educational or signing courses.

User 4:CP

Female aged 26 at date of participation

British Picture Vocabulary Scale (Short Form)

Raw Score:	9
Age Equivalent:	4 years 0 months
Confidence Interval:	3-4 to 4-8

U4 is Down syndrome with both visual and hearing impairment. However, both are corrected to within the *normal* range with glasses and hearing aids. She is classified as having moderate learning handicap. Despite her learning difficulties U4's adaptive, or functional behaviour, is extremely good and she lives alone in a flat as part of a sheltered housing complex. Her monetary skills are good for example, as a result of extensive training courses which she continues to attend. She has also attended courses for rudimentary computer work, and can easily type her name and address etc...

Her speech is poor however, probably as a result of deafness, with slurring common. Combined with low production rates, and rapid bursts intelligibility is poor. She has a rudimentary knowledge of Makaton so as to be able to interact with her friends.

Her memory and attention span are good as one might expect from her levels of adaptive skill. She is eager to learn any new material and will persevere until she has leant it to the best of her ability.

Her motor skills are good and she enjoys a variety of sports.

Justification for using borderline verbal LD users

Although Easy Speaker was designed to be used by the non, or barely vocal, LD population as a primary method of communication it was decided that for initial evaluative work this would be impractical. Consequently, users were chosen from an available population of LD users who were vocal, but who had average to severe language problems. All however could communicate to a greater or lesser extent through their caregivers, and with time, directly with the author. This meant that they could inform the author as to what they thought of Easy Speaker and the strategies they were using to control it, albeit in an unsophisticated manner. Due to the limits of their ability this fell short of full protocol analysis but was highly informative none-the-less and helped to influence the design of extra features, and the modification of others. All were considered to possess the ability to use Easy Speaker as their primary means of communication if necessitated.

In addition, as the author only had six months available to field trial Easy Speaker it was felt that withdrawal of support for more severely handicapped individuals would have been unethical. All users in the current study understood that they may be able to continue to use Easy Speaker after the six months trial, but without the intense support of the author. To this end intervention was steered toward Easy Speaker and did not focus around the visits of the author.

Apparatus

An ICL 386 SL notebook computer with external VGA colour monitor and keyboard. A LogiTech Audioman connected to the computers parallel port was used for digitised speech sampling and reproduction. A standard mouse was also used.

A copy of the "Easy Speaker Assessment Tasks" (ESAT) developed by the author was utilised in all assessments. See chapter 8 for a full discussion of development, administration and scoring.

Procedure

At week five of the main 24 week Easy Speaker study users were introduced to the Easy Speaker Assessment Tasks (ESAT). The preceding four week familiarisation period was designed to enable users to acclimatise to the author, the computer equipment, and moreover to learn the rudiments of controlling and using Easy Speaker. An overview of how the ESAT tasks slotted into the main study is shown below :-

Fig 9.2, How ESAT tasks related to the main body of the study

Week No./ session No. (timeframe)	Timeframe description	Weekly sessions	Hours per session	Sessions in timeframe	Total hours in timeframe
1-4	Familiarisation/acclimatisation period	1	2	4	8
5	ESAT tasks week 1	1	2	1	2
6	ESAT tasks week 2	1	2	1	2
7	ESAT tasks week 3	1	2	1	2
8-24	Main body of full study	1	2	17	34

Total number of hours per user over the 24 week study period	48
Total number of hours devoted to all users during each week	8
Total number of hours devoted to users as a whole over the 24 week study period	192

During the four week familiarisation period (weeks 1-4) all users learned to control Easy Speaker cursor movement by means of the four cursor [arrowed] keys embedded within the standard 102 key IBM PC keyboard, to use the return key to select the current speech symbol or link, and to use the space bar to reproduce the current utterance. Users were prompted to explore any parts of the hierarchy they wished, but were not prompted to, or barred from exploring those levels and speech symbols which would be used within ESAT tasks. In total each user had eight hours, spread over four weeks, in which to learn how to use Easy Speaker to the best of their ability. Any queries and/or reservations which users, or caregivers had were also resolved within this period. The base hierarchy which all used consisted of around 1200 symbols organised according to logical categories with appropriate links between them. Each symbol represented a variable length of digitised speech, e.g. from anywhere between one word to seven making up a full sentence.

Over the next three weeks (weeks 5-7) users were evaluated using the ESAT tasks. That is, each user was required to complete each set of tasks once on each of the three weekly sessions. Within each of the three sessions users were allowed the first 20 minutes or so to re-orientate themselves with Easy Speaker and its controls. This simply took the form of free/semi-structured use of Easy Speaker before being asked to construct and generate the eight ESAT task utterances as described in chapter eight.

Selection of Performance Criteria

Although Easy Speakers in-built tracking system, ARTS, records virtually all aspects of a users interaction, for the purposes of measuring performance gains I have found that three key measures are sufficient. Two of the measures are recorded verbatim, with the third being calculated by the tracking system.

The first of the two actual measures is known as total time (TOT T) and is a measure, in seconds and hundredths, of the elapsed time between starting to construct an utterance and the time when it is generated as speech. A decreasing TOT T would suggest an increase in the speed with which the user could navigate the hierarchy and generate speech. This might be explained purely as a speed improvement, by improvements in navigational efficiency, or by a combination of the two.

The second is known as Total Mickey Count (TMC) and is a measure of how far in mickeys (units of approx. 1 mm on screen) the cursor has moved on the desktop between starting to construct the utterance and the time it is generated as speech. The mickey count can simply be thought of as an 'odometer-like' measure that records how far the cursor has travelled in constructing each utterance. As one would hope that users develop more efficient navigational strategies overtime, TMC is a good measure of this improvement. A decreasing TMC would suggest an increase in navigational efficiency, or mouse tack. It is worth pointing out the although all users in this study controlled the Easy Speaker cursor with the cursor keys, using a measure such as mickeys is still valid as the keys directly emulated the mouse at a hardware level (in the Windows messaging loop to be precise).

The final calculated measure is an estimate of how many Words Per Minute (WPM) the user would be likely to construct and generate based on their TOT T for the last utterance and the number of words it contained. Although Easy Speaker uses digitised speech the information file for each symbol contains a textual transcription of the digitised speech it represents and the number of words it contains. Thus an approximate words per minute score can be computed. The

Words Per Minute estimate varies with each utterance constructed and is calculated internally using the following formula :-

WPM = 60 sec's / [TOT T / number of words in utterance]

For example, to calculate the words per minute score for "get the nurse - sore - back" from the sample norms assuming a user took 137.58 seconds to construct it :-

60 / [137.58 / 5] = 2.18 WPM

The estimated WPM simply provides a 'rough and ready' measure of users ability and the potential of Easy Speaker itself. One must bear in mind that as any speech symbol could represent several sentences of speech and a relatively high, or low, WPM might be misleading when viewed in this context.

In addition to purely performance measures qualitative assessment of general output was also made, e.g. analysis for evidence of requests and statements made on device.

Calculation of Sample Norms and Best Case Norms

Although analyses concentrates on individual users performance it is useful to apply two sets of calculated norms. For example, in addition to knowing whether a users performance is improving over sessions it is useful to know whether a user is performing significantly above average, or to know whether a user is regressing in performance terms toward the performance achieved by an unimpaired user etc... These norms are known as Sample Norms and Best Case Norms respectively. Sample Norms reflect the average sample performance on the three criterion measures for the ESAT tasks. Best Case Norms represent an unimpaired users average performance on the three criterion measures for the ESAT tasks.

Sample norms were calculated retrospectively by averaging scores for all users across all ESAT tasks for all three sessions. For example, to find the sample norm for the total time taken (TOT T) to construct the utterance "meals" the following method was used :-

1. Total all TOT T data for "meals" for all users and all sessions together, e.g.

   54.49+28.84+..+27.46+21.75 = 508.82 seconds

2. Total number of times "meals" was constructed, e.g.

   No. of users * Total no. of sessions per user = 4 * 3 = 12

3. Divide total TOT T by total number of times "meals" was constructed to get the sample norm, e.g.

   508.82/12 = 42.40 sec's

The actual sample norms calculated from the four users performances on each of the three ESAT sessions are shown below in table form :-

Fig 9.3, Sample norms for total time (TOT T), total mickey count (TMC) and words per minute (WPM)

Sample Norms

Task No.	Transcription of utterence subjects had to construct	TOT T	TMC	WPM
1	meals	42.40	130.67	1.65
2	summer	65.09	180.17	1.17
3	trainers	74.83	180.75	0.95
4	muscles	110.26	221.50	0.62
5	ham/salad	88.29	251.58	1.57
6	read/library/book	116.52	284.25	1.84
7	is it/raining	121.15	358.92	1.72
8	get the nurse/sore/back	137.58	357.83	2.59
	Mean	94.51	245.71	1.51
	SD	32.30	83.87	0.61

Key: TOT T: Total time in seconds, TMC: Total mickey count in mickeys, WPM: Words per minute (estimated)

The second set of norms that were used are best case norms. Best case norms were calculated by averaging an experienced unimpaired users ESAT task scores over three sessions. The actual best case norms are shown below in table form :-

Fig 9.4, Best case norms for total time (TOT T), total mickey count (TMC) and words per minute (WPM)

Task No.	Transcription of utterence subject had to construct	Best Case Norms TOT T	TMC	WPM
1	meals	7.30	126.33	8.22
2	summer	7.45	85.33	8.05
3	trainers	12.21	87.67	4.91
4	muscles	16.26	122.33	3.69
5	ham/salad	13.14	142.00	9.13
6	read/library/book	14.74	122.33	16.28
7	is it/raining	16.49	143.67	10.92
8	get the nurse/sore/back	17.28	161.67	17.36
	Mean	13.11	123.92	9.82
	SD	3.93	26.61	4.89

Key: TOT T: Total time in seconds, TMC: Total mickey count in mickeys, WPM: Words per minute (estimated)

Overview of the Presentation of Results and Discussion

In order to present the complex data generated by the ARTS, and statistical analyses carried out by the author in a coherent way, results are presented for individual users separately. The results presented for U1 will be explained in depth and shall serve a model for the remaining three users who's results only will be presented.

As each users results are presented discussion of those results and suggestions regarding possible implications are made. Once all results have been presented and discussed in turn, a general discussion will consider the findings as a whole.

Results and Discussion for Individual Users

Results for User 1: PS

Descriptive statistical analysis suggested that U1's performance improved on all measures with each passing session. U1's actual performance data for the three sessions is shown below in table form :-

Fig 9.5, Descriptive statistics for U1's ESAT sessions

Task No.	Transcription of utterence subject had to construct	Session 1 TOT T	Session 1 TMC	Session 1 WPM	Session 2 TOT T	Session 2 TMC	Session 2 WPM	Session 3 TOT T	Session 3 TMC	Session 3 WPM
1	meals	54.49	138.00	1.10	28.84	101.00	2.08	25.81	78.00	2.32
2	summer	87.60	118.00	0.68	36.53	101.00	1.64	41.46	83.00	1.45
3	trainers	73.99	149.00	0.81	57.06	106.00	1.05	47.01	94.00	1.28
4	muscles	116.72	165.00	0.51	112.21	138.00	0.53	91.34	130.00	0.66
5	ham/salad	104.09	204.00	1.15	61.85	143.00	1.94	67.61	131.00	1.77
6	read/library/book	144.18	198.00	1.25	86.73	158.00	2.08	73.66	136.00	2.44
7	is it/raining	117.82	236.00	1.53	83.98	139.00	2.14	88.04	147.00	2.04
8	get the nurse/sore/back	216.74	234.00	1.38	115.95	183.00	2.59	111.77	190.00	2.68
	Mean	114.45	180.25	1.05	72.89	133.63	1.76	68.34	123.63	1.83
	SD	49.89	44.31	0.35	32.40	29.42	0.66	28.83	37.44	0.68

Key: TOT T: Total time in seconds, TMC: Total mickey count in mickeys (1 mickey = approx 1 mm) , WPM: Words per minute (estimated)

In performance terms utterance construction[1] times [TOT T] speeded up, over each of the three sessions (M (8) session 1 = 114.45 sec's; M (8) session 2 = 72.89 sec's; M (8) session 3 = 68.34 sec's).

Efficiency of cursor movement[2], or tack [TMC], U1 became more efficient over the three sessions (M (8) session 1 = 180.25 mickeys; M (8) session 2 = 133.63 mickeys; M (8) session 3 = 123.63 mickeys).

As one would expect the estimated number of words per minute[3] [WPM] also improves over the three sessions (M (8) session 1 = 1.05 wpm; M (8) session 2 = 1.76 wpm; M (8) session 3 = 1.83 wpm).

In order to examine these improvements further it is useful to breakdown the analyses into the three main criterion measures :-

1. TOTal Time taken	TOT T
2. Total Mickey Count	TMC
3. Words Per Minute (estimated)	WPM

1 Construction in this context refers to construction and generation

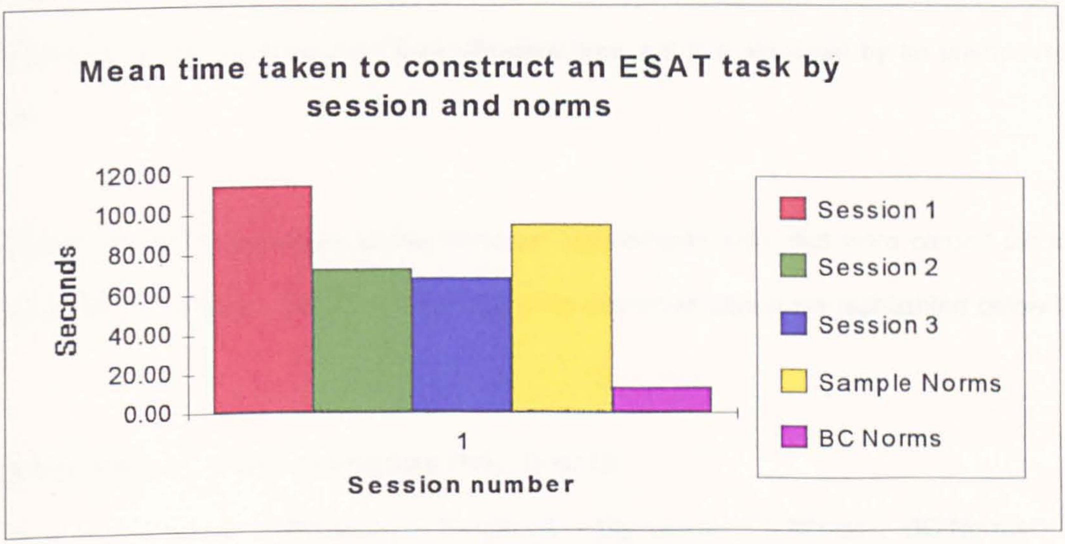
2 Measured in Mickey's: 1 Mickey = approximately 1 mm on the screen desktop

3 Estimated from TOT T and the number of words in utterance

Total time taken (TOT T)

To obtain a 'course grain' picture of the performance improvements in utterance construction times made by U1 over the sessions it is useful to plot the means for each session along with sample and best case norms :-

Fig 9.6, Mean time (TOT T) taken to construct an ESAT task by session and norms for U1



Although U1's performance improvements are clearly illustrated the question of whether these are significant improvements must be addressed.

As all ESAT data is non-parametric in nature and relates to repeated measures several Wilcoxon signed-ranks tests were applied to assess the significance of these improvements. The first focused on whether U1 significantly improved in construction time as compared with the first session. These suggested there was a significant improvement between session 1 and session 2 ($Z = -2.52$, sig. 0.01), and session 1 and session 3 ($Z = -2.52$, sig. 0.01).

The second focused on whether U1 significantly improved in construction time with each successive session. This suggested there was a significant improvement between session 1 and session 2 ($Z = -2.52$, sig. 0.01). However there was no significant improvement between session 2 and session 3 ($Z = -0.98$, NS).

The third focused on whether U1 significantly improved in construction time on his third session as compared with the sample norms. This suggested there was a significant improvement beyond the sample norm (Z = -2.38, sig. 0.01).

The fourth focused on whether U1 significantly improved in construction time on his third session as compared with the best case norms. This suggested that U1 had not improved so significantly as to approach the construction time of the best case norm (Z = -2.52, sig 0.01), i.e. there was still a significant difference between U1's construction time and that achieved by an unimpaired individual.

The following table summarises all the Wilcoxon signed-ranks tests that were carried out in relation to construction time. The Wilcoxon (Z) results described above are highlighted below in bold :-

Fig 9.7, Wilcoxon Z matrix for total time (TOT T) for U1

	Session 1	Session 2	Session 3	S Norms	BC Norms
Session 1	-	-	-	-	-
Session 2	**-2.52****	-	-	-	-
Session 3	**-2.52****	**-0.98**	-	-	-
S Norms	-2.10*	-2.38**	**-2.52****	-	-
BC Norms	-2.52**	-2.52**	**-2.52****	-2.52**	-

Key to significance levels: * 0.05, ** 0.01

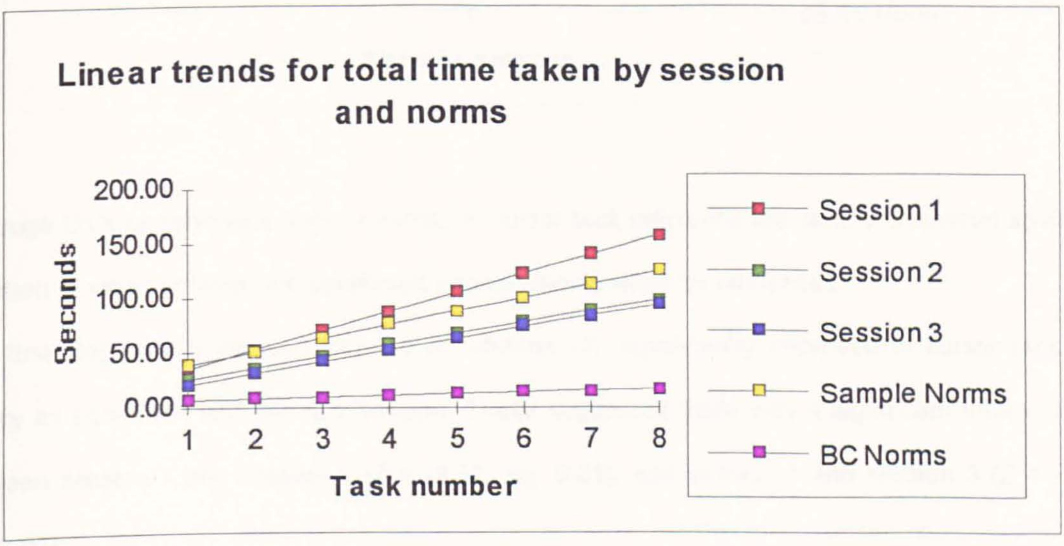
To confirm there was a significant improvement in total time over the three sessions a Friedman test was carried out yielding a Chi square value of 12.25 with 2 df which is significant at the 0.005 level. This suggests that this pattern of improvement in construction time is highly unlikely to be the result of chance.

In order to obtain a 'finer grain' picture of performance improvements in construction times made by U1 over the three sessions it is useful to calculate linear trends for each session using U1's actual performance data and plot these along with sample and best case norms :-

Fig 9.8, A summary of slopes and intercepts for total time (TOT T) for U1

	Slope	Intercept
Session 1	17.68	34.91
Session 2	10.54	25.45
Session 3	10.60	20.61
S Norms	12.50	38.28
BC Norms	1.42	6.71

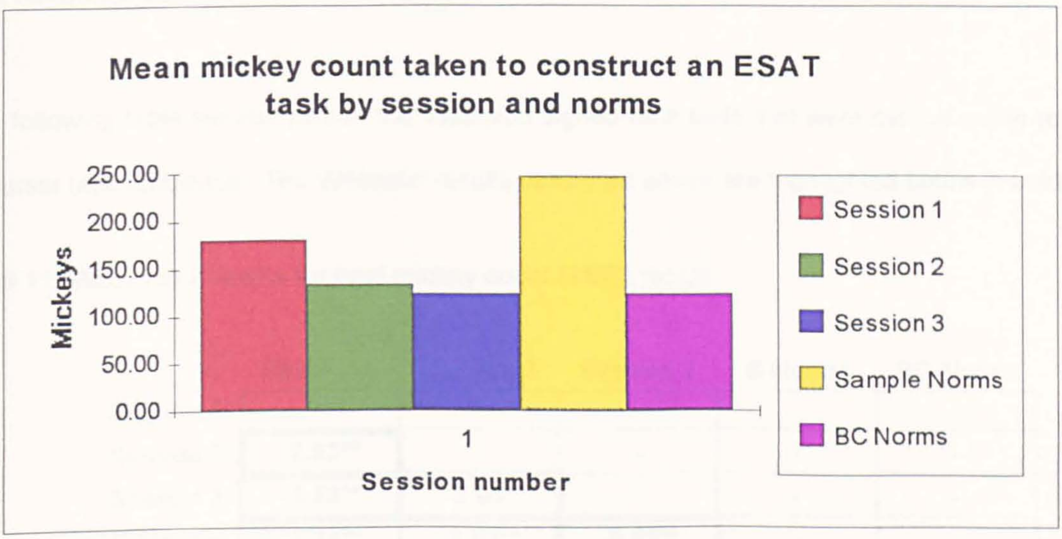
Fig 9.9, Calculated total time (TOT T) linear trends for each session and norms U1



Total mickey count (TMC)

To obtain a 'course grain' picture of the performance improvements in cursor tack efficiency made by U1 over the sessions it is useful to plot the means for each session along with sample and best case norms :-

Fig 9.10, Mean total mickey count (TMC) taken to construct all utterances by session and norms for U1



Although U1's performance improvements in cursor tack efficiency are clearly illustrated again the question of whether these are significant improvements must be addressed.

The first statistical evaluation focused on whether U1 significantly improved in cursor tack efficiency as compared with the first session. These suggested there was a significant improvement between session 1 and session 2 (Z = -2.52, sig. 0.01), and session 1 and session 3 (Z = -2.52, sig. 0.01).

The second focused on whether U1 significantly improved in cursor tack efficiency with each session. These suggested there was a significant improvement between session 1 and session 2 (Z = 2.52, sig. 0.01), and between session 2 and session 3 (Z = 2.03, Sig. 0.05).

The third focused on whether U1 significantly improved in cursor tack efficiency on his third session as compared with the sample norms. This suggested there was a significant improvement beyond the sample norm (Z = 2.52, sig. 0.01).

The fourth focused on whether U1 significantly improved in cursor tack efficiency on his third session as compared with the best case norm. This suggested that U1 had improved so significantly as reach the cursor tack efficiency of the best case norm (Z = -0.56, NS), i.e. there was no significant difference between U1's cursor tack efficiency and that of an unimpaired user.

The following table summarises all the Wilcoxon signed-rank tests that were carried out in relation to cursor tack efficiency. The Wilcoxon results described above are highlighted below in bold :-

Fig 9.11, Wilcoxon Z matrix for total mickey count (TMC) for U1

	Session 1	Session 2	Session 3	S Norms	BC Norms
Session 1	-	-	-	-	-
Session 2	**-2.52****	-	-	-	-
Session 3	**-2.52****	**-2.03***	-	-	-
S Norms	-2.38**	-2.52**	**-2.52****	-	-
BC Norms	-2.52**	-1.26	**-0.56**	-2.52**	-

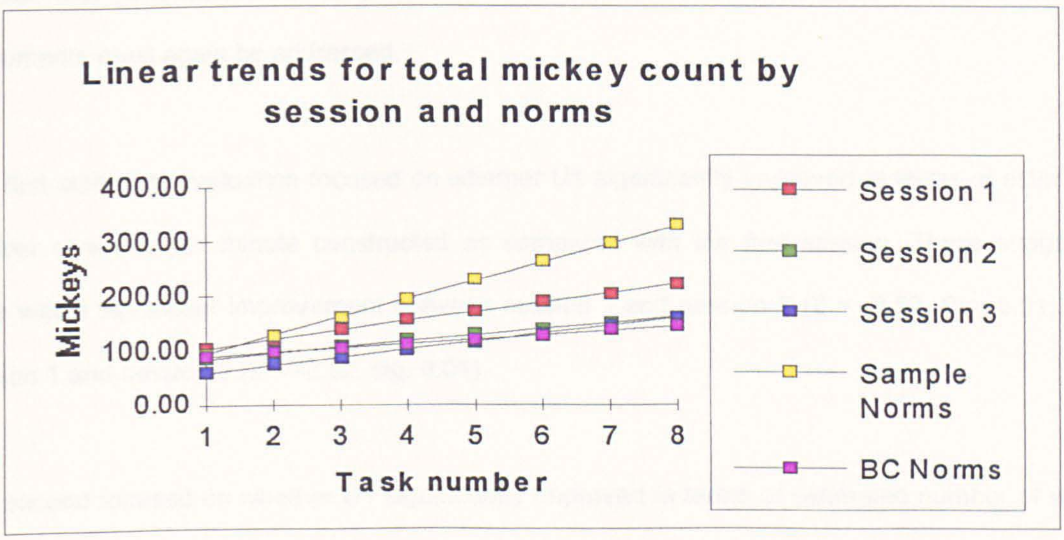
Key to significance levels: * 0.05, ** 0.01

To confirm there was a significant improvement over the three sessions a Friedman test was carried out yielding a Chi square value of 13.00 with 2 df, which is significant at the 0.001 level. Again it is useful to calculate linear trends for each session using the data from U1's actual performance data and plot these along with sample and best case norms :-

Fig 9.12, A summary of slopes and intercepts for total mickey count (TMC) for U1

	Slope	Intercept
Session 1	17.23	102.68
Session 2	11.01	84.07
Session 3	14.65	57.67
S Norms	33.62	94.40
BC Norms	7.89	88.41

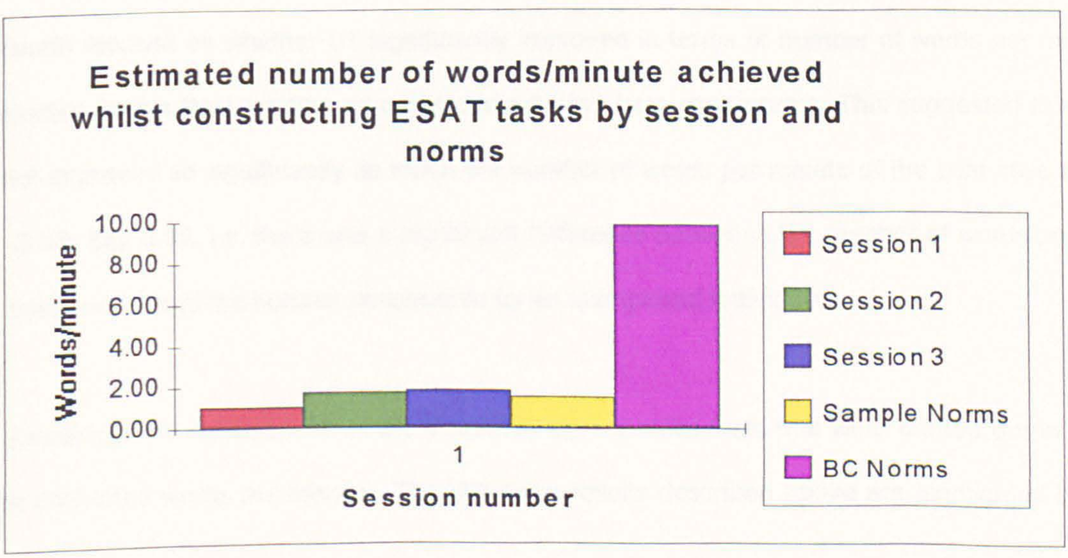
Fig 9.13, Calculated total mickey count (TMC) linear trends for each session and norms for U1



Words per minute (WPM)

To obtain a 'course grain' picture of the performance improvements in estimated words per minute constructed by U1 over the three sessions it is useful to plot the means for each session along with sample and best case norms :-

Fig 9.14, Mean number of estimated words per minute (WPM) constructed by session and norms for U1



Although U1's performance improvements in the estimated number of words per minute constructed and generated are clearly illustrated the question of whether these are significant improvements must again be addressed.

The first statistical evaluation focused on whether U1 significantly improved in terms of estimated number of words per minute constructed as compared with the first session. These suggested there was a significant improvement between session 1 and session 2 (Z = -2.52, Sig. 0.01), and session 1 and session 3 (Z = -2.52, sig. 0.01).

The second focused on whether U1 significantly improved in terms of estimated number of words per minute constructed with each session. These suggested there was a significant improvement between session 1 and session 2 (Z = -2.52, sig. 0.01). However between session 2 and session 3 there was no significant improvement (Z = -0.98, NS).

The third focused on whether U1 significantly improved in terms of estimated number of words per minute constructed on his third session as compared with the sample norms. This suggested there was a significant improvement beyond the sample norm (Z = -2.52, sig. 0.01).

The fourth focused on whether U1 significantly improved in terms of number of words per minute constructed on his third session as compared with the best case norms. This suggested that U1 had not improved so significantly as reach the number of words per minute of the best case norm (Z = -2.52, Sig. 0.01, i.e. there was a significant difference between U1's number of words per minute constructed and the number constructed by an unimpaired individual.

The following table summarises all the Wilcoxon signed-ranks tests that were carried out in relation to estimated words per minute. The Wilcoxon results described above are highlighted below in bold :-

Fig 9.15, Wilcoxon Z matrix for estimated number of words per minute (WPM) constructed for U1

	Session 1	Session 2	Session 3	S Norms	BC Norms
Session 1	-	-	-	-	-
Session 2	**-2.52****	-	-	-	-
Session 3	**-2.52****	**-0.98**	-	-	-
S Norms	-1.96*	-2.38**	**-2.52****	-	-
BC Norms	-2.52**	-2.52**	**-2.52****	-2.52**	-

Key to significance levels: * 0.05, ** 0.01

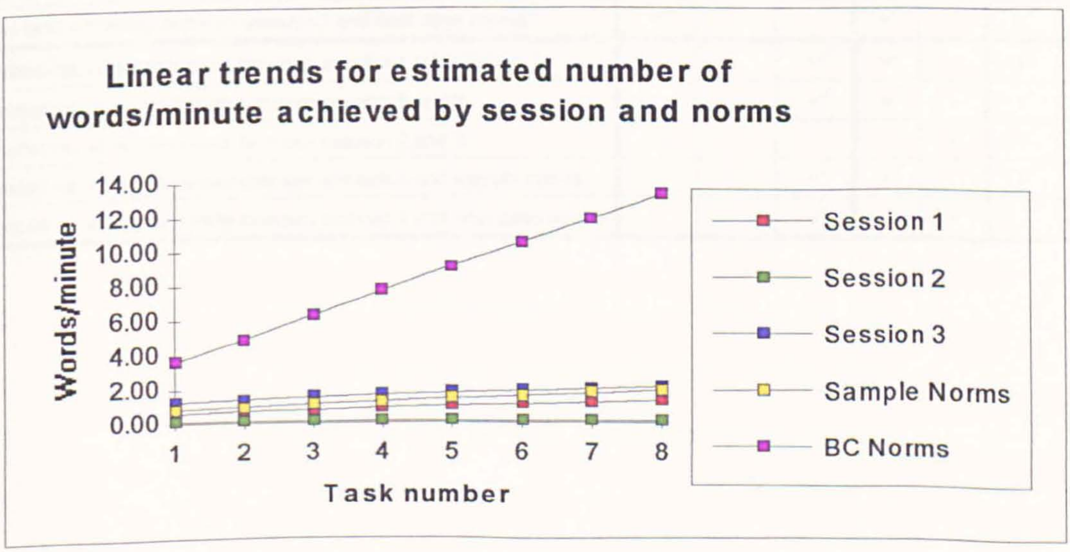
To confirm there was a significant improvement over the three sessions a Friedman test was carried out yielding a Chi square value of 12.25 with 2 df which is significant at the 0.005 level.

Again it is useful to calculate linear trends for each session using the data from U1 actual performance data and plot these along with sample and best case norms :-

Fig 9.16, A summary of slopes and intercepts for estimated number of words per minute constructed (WPM) for U1

	Slope	Intercept
Session 1	0.10	0.61
Session 2	0.13	1.19
Session 3	0.12	1.29
S Norms	0.13	0.68
BC Norms	1.26	3.65

Fig 9.17, Estimated number of words per minute (WPM) constructed linear trends for each session and norms for U1



One can summarise ESAT task results for U1 by means of the following table. Significance levels are taken directly from Wilcoxon signed-ranks tests presented earlier. The "Significant Improvement" columns differentiate whether the significance level quoted denotes a significant improvement or not in the users level of performance. For example, there is no significant (NS) difference between "Cursor tack efficiency between session 3 and best case norms". However this is an improvement as U1 has reached the performance of an unimpaired user and so is signified by a tick in the "Yes" side of the "Improvement" column. The "Improvement" columns denote whether there has been an improvement or not regardless of whether that improvement has achieved significance. It must be remembered that within such an LD group any improvement is personally significant regardless of whether it achieves statistical significance or not.

Fig 9.18, Summary of ESAT performance improvements for U1

	Significance Level			Significant Improvement		Improvement	
Differences in ...	NS	0.05	0.01	Yes	No	Yes	No
Speed between session 1 and 2			✓	✓		✓	
Speed between session 1 and 3			✓	✓		✓	
Speed between session 2 and 3	✓				✓	✓	
Speed between session 3 and sample norms			✓	✓		✓	
Speed between session 3 and best case norms			✓		✓		✓
Cursor tack efficiency between session 1 and session 2			✓	✓		✓	
Cursor tack efficiency between session 1 and 3			✓	✓		✓	
Cursor tack efficiency between session 2 and 3		✓			✓	✓	
Cursor tack efficiency between session 3 and sample norms			✓	✓		✓	
Cursor tack efficiency between session 3 and best case norms	✓			✓		✓	
Estimated no. of words/minute between session 1 and session 2			✓	✓		✓	
Estimated no. of words/minute between session 1 and 3			✓	✓		✓	
Estimated no. of words/minute between session 2 and 3	✓				✓	✓	
Estimated no. of words/minute between session 3 and sample norms			✓	✓		✓	
Estimated no. of words/minute between session 3 and best case norms			✓		✓		✓

DISCUSSION FOR USER 1: PS

U1 had little trouble in completing all the EAST tasks presented, and only required the bare minimum of help when first attempting them. Despite U1's relatively poor BPVS and other test scores, his memory functionality remains intact and helps account for his good performance.

In terms of the mean time taken to construct an ESAT task, U1 improved with each session. Construction time improved markedly with a mean drop of 41.56 seconds between session 1 and session 2. However by session 3 improvements were down to a mean drop of 4.55 seconds over session 2. This suggests that improvements in construction speed would continue to take place but according to the law of diminishing returns. U1 made very significant improvements in construction time over the three sessions as a whole.

In efficiency terms U1, against all expectations, by the third session had reached the level of efficiency shown by an unimpaired individual for the same tasks. His levels of efficiency improvement between the three sessions were extremely significant as one might expect. This is even more remarkable when one considers that U1 used a potentially inefficient cursor movement strategy to begin with. That is, instead of keeping his finger on a cursor key to keep the cursor moving in the required direction he would tap the key repeatedly at speed to achieve the same effect. However, this technique may have allowed him to constantly adjust his cursor tack with each key press. Whereas, unimpaired individuals generally keep their finger down on the key until the cursor nears the point where they should change tack or they have reached their target. In some cases this 'pre-plotting' of cursor tack combined with high cursor movement speed leads them to overshoot the desired target. It is worth noting that U1 never overshot the target using this strategy and in effect honed his efficiency. His strategy was totally natural as no users were prompted or dissuaded from using any technique they liked in using Easy Speaker, unless they were obviously disadvantageous. It is worth empathising that constructional speed and efficiency need not be linked. Simply, a user may be slow yet extremely efficient. Conversely, they may be very quick but inefficient. One virtue can compensate for the other weakness. In U1's case his high level of efficiency, despite his slow careful planning, led to above average constructional speed overall.

Using such a strategy however can have negative consequences. The first being that the speed with which the cursor can be moved is restricted to how quickly the user can tap the key. Whereas the 'auto-repeat' built into the keyboard is generically faster. Secondly repeated key tapping is a very poor ergonomic strategy and could potentially lead to Repetitive Strain Injury (RSI). Obviously a touch screen would eliminate these difficulties and increase constructional speed dramatically and lead to increased efficiency.

When one examines U1's construction speed and his high levels of efficiency more closely one can propose that these were a direct result of him *learning the hierarchy*. Given his apparently unnoticed, but excellent memory, it is easy to see why this might be the case. This in turn suggests that U1 has a good capacity for learning given an increased level of motivation and a suitable approach. In this case what might be considered a multi-sensory one.

In addition to U1's learning of the hierarchy, his efficiency suggests that he formulated and internalised an effective strategy for navigating the cursor around the hierarchy. That is, U1 set navigational goals, planned how to achieve those goals effectively, and achieved them efficiently. Indeed when asked U1 could verbally describe the strategy he had used for a preceding task. This suggests, that his metacognitive functioning is also above expected levels. One might even suggest that such enriched exposure to goal orientated tasks would be likely to enhance cognitive, or metacognitive functioning (reflectivity).

Occasional observers who knew the user well also noted how his level of on task attention was increased substantially above normal. His levels of impulsively were also commented on as they also were observably below the norm. For these occasional observers, the most striking effect of being embroiled in using Easy Speaker was that U1's normal level of 'finger flicking' and other stereotyped behaviours were vastly reduced. One might attribute this reduction to his increased task attention and lowered general levels of anxiety resulting from on-task focusing. Contrary to popular opinion U1 never displayed any sign of the severe Challenging Behaviour (CB) for which he was renowned, e.g. spontaneous head-butting etc..., whilst using Easy Speaker, or in the presence of the author. This might be explained in two ways. Firstly that U1 was so task focused that the CB was virtually extinguished. Secondly, that U1's CB was an outlet used to gain attention in an otherwise boring institutionalised environment.

As U1's speech was lacking in quality and constructional complexity it was pleasing to note that he would mimic, and on occasion, pre-empt Easy Speakers spoken output by speaking the sentence or phrase he had just assembled.

USER 2: AW

Fig 9.19, Descriptive statistics for U2's ESAT sessions

		Session 1			Session 2			Session 3		
Task No.	Transcription of utterence to construct and generate	TOT T	TMC	WMIN	TOT T	TMC	WMIN	TOT T	TMC	WMIN
1	meals	60.42	192.00	0.99	50.42	189.00	1.19	44.60	96.00	1.35
2	summer	62.61	123.00	0.96	58.23	147.00	1.03	59.15	164.00	1.01
3	trainers	65.52	113.00	0.92	61.57	106.00	0.97	55.96	110.00	1.07
4	muscles	89.25	157.00	0.67	94.03	160.00	0.64	110.18	150.00	0.54
5	ham/salad	98.48	450.00	1.22	104.97	208.00	1.14	82.83	189.00	1.45
6	read/library/book	106.89	200.00	1.68	94.15	181.00	1.91	90.14	151.00	2.00
7	is it/raining	132.26	274.00	1.36	118.75	330.00	1.52	130.18	218.00	1.38
8	get the nurse/sore/back	124.96	257.00	2.40	132.70	286.00	2.26	129.90	241.00	2.31
	Mean	92.55	220.75	1.28	89.35	200.88	1.33	87.87	164.88	1.39
	SD	28.13	109.02	0.55	29.97	73.76	0.53	33.38	49.81	0.56

Total time (TOT T)

Fig 9.20, Mean time (TOT T) taken to construct an ESAT task by session and norms for U2

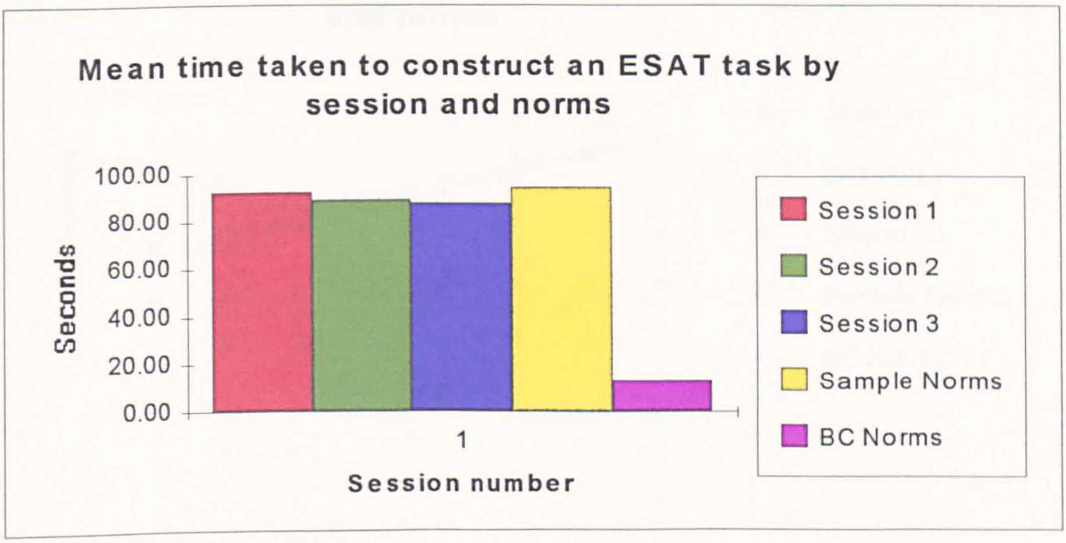


Fig 9.21, Wilcoxon Z matrix for total time (TOT T) for U2

	Session 1	Session 2	Session 3	S Norms	BC Norms
Session 1	-	-	-	-	-
Session 2	**-0.84**	-	-	-	-
Session 3	**-0.98**	**-0.56**	-	-	-
S Norms	-0.28	-0.98	**-1.40**	-	-
BC Norms	-2.52**	-2.52**	**-2.52****	-2.52**	-

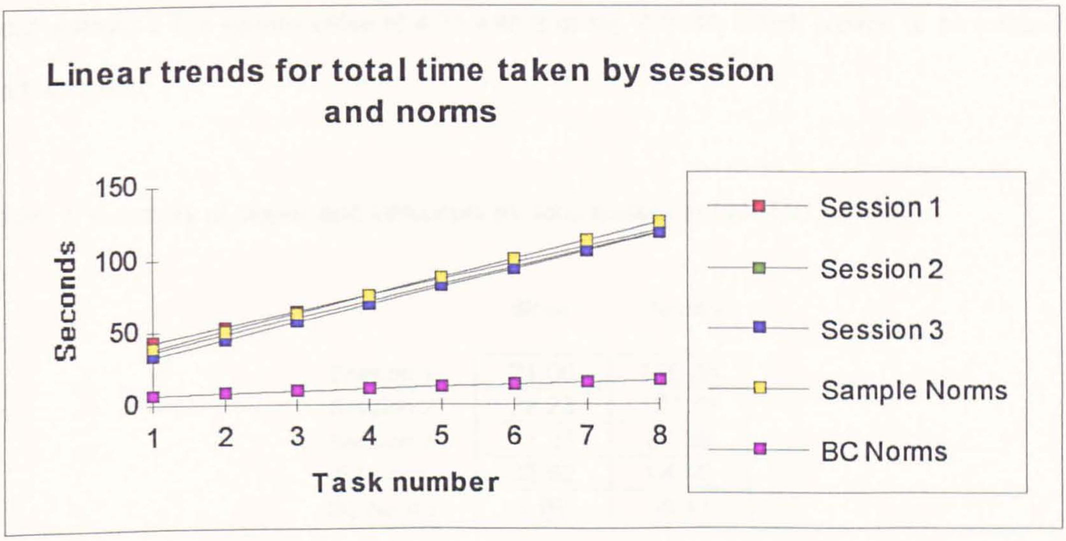
Key to significance levels: * 0.05, ** 0.01

To confirm there was a significant improvement over the three sessions a Friedman test was carried out yielding a Chi square value of 2.25 with 2 df sig. 0.3247, which proved to be insignificant at the 0.05 level.

Fig 9.22, A summary of slopes and intercepts for total time (TOT T) for U2

	Slope	Intercept
Session 1	11.11	42.55
Session 2	11.75	36.46
Session 3	12.83	32.83
S Norms	12.50	38.28
BC Norms	1.42	6.71

Fig 9.23, Calculated total time (TOT T) linear trends for each session and norms for U2



Mean total mickey count (TMC)

Fig 9.24, Mean total mickey count (TMC) taken to construct an ESAT task by session and norms for U2

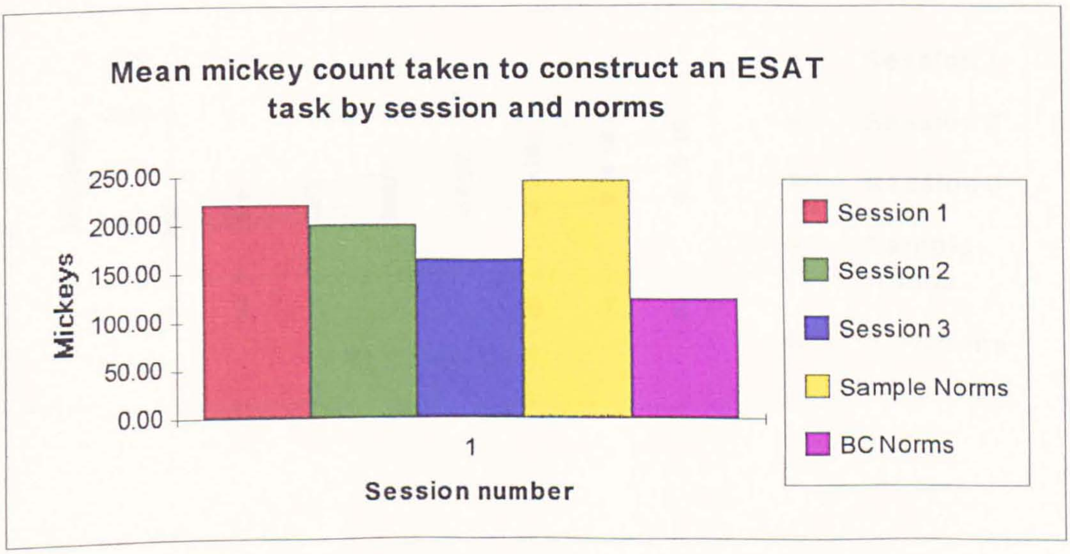


Fig 9.25, Wilcoxon Z matrix for mickey count (TMC) for U2

	Session 1	Session 2	Session 3	S Norms	BC Norms
Session 1	-	-	-	-	-
Session 2	**-0.21**	-	-	-	-
Session 3	**-1.96***	**-1.96***	-	-	-
S Norms	-1.12	-1.96*	**-2.52****	-	-
BC Norms	-2.52**	-2.52	**-1.96***		-

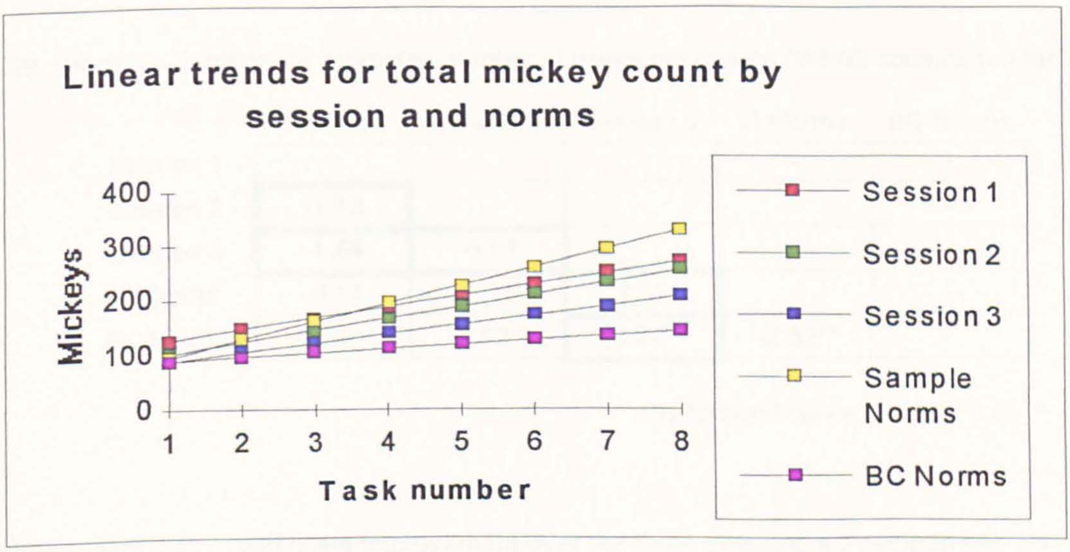
Key to significance levels: * 0.05, ** 0.01

To confirm there was a significant improvement over the three sessions a Friedman test was carried out yielding a Chi square value of 4.75 with 2 df sig. 0.0930, which proved to be insignificant at the 0.05 level.

Fig 9.26, A summary of slopes and intercepts for total mickey count (TMC) for U2

	Slope	Intercept
Session 1	21.00	126.25
Session 2	22.23	100.86
Session 3	17.23	87.36
S Norms	33.62	94.40
BC Norms	7.89	88.41

Fig 9.27, Calculated total mickey count (TMC) linear trends for each session and norms for U2



Words per minute (WPM)

Fig 9.28, Mean number of estimated words per minute (WPM) constructed by session and norms for U2

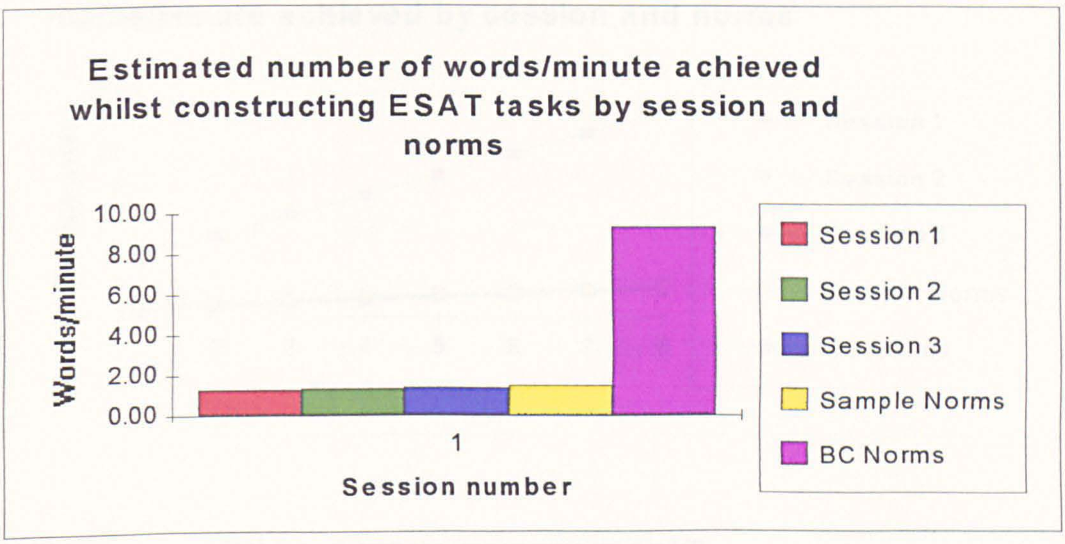


Fig 9.29, Wilcoxon Z matrix for estimated number of words per minute (WPM) constructed for U2

	Session 1	Session 2	Session 3	S Norms	BC Norms
Session 1	-	-	-	-	-
Session 2	**-1.12**	-	-	-	-
Session 3	**-1.54**	**-0.91**	-	-	-
S Norms	-0.14	-0.70	**1.52**	-	-
BC Norms	-2.52**	-2.52**	**-2.52****	-2.52**	-

Key to significance levels: * 0.05, ** 0.01

To confirm there was a significant improvement over the three sessions a Friedman test was carried out yielding a Chi square value of 2.25 with 2 df sig. 0.3247, which proved to be insignificant at the .05 level.

Fig 9.30, A summary of slopes and intercepts for estimated number of words per minute constructed (WPM) for U2

	Slope	Intercept
Session 1	0.18	0.49
Session 2	0.16	0.62
Session 3	0.15	0.73
S Norms	0.13	0.68
BC Norms	1.26	3.65

Fig 9.31, Estimated number of words per minute (WPM) constructed linear trends for each session and norms for U2

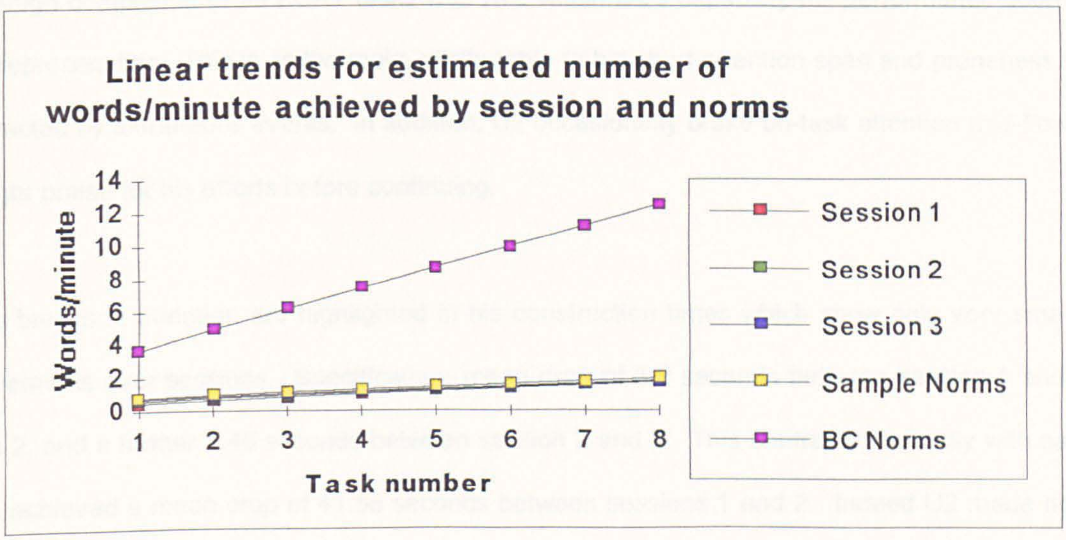


Fig 9.32, Summary of ESAT performance improvements for U2

Differences in ...	Significance Level NS	0.05	0.01	Significant Improvement Yes	No	Improvement Yes	No
Speed between session 1 and 2	✓				✓	✓	
Speed between session 1 and 3	✓				✓	✓	
Speed between session 2 and 3	✓				✓	✓	
Speed between session 3 and sample norms	✓				✓	✓	
Speed between session 3 and best case norms			✓		✓		✓
Cursor tack efficiency between session 1 and session 2	✓				✓	✓	
Cursor tack efficiency between session 1 and 3		✓		✓		✓	
Cursor tack efficiency between session 2 and 3		✓		✓		✓	
Cursor tack efficiency between session 3 and sample norms			✓	✓		✓	
Cursor tack efficiency between session 3 and best case norms		✓			✓		✓
Estimated no. of words/minute between session 1 and session 2	✓				✓	✓	
Estimated no. of words/minute between session 1 and 3	✓				✓	✓	
Estimated no. of words/minute between session 2 and 3	✓				✓	✓	
Estimated no. of words/minute between session 3 and sample norms	✓				✓		✓
Estimated no. of words/minute between session 3 and best case norms			✓		✓		✓

DISCUSSION FOR USER 2: AW

Although completion of all ESAT tasks was well within U2's capability his performance scores do not represent this. This is, in the main, attributable to his short attention span and proneness to be distracted by extraneous events. In addition, U2 occasionally broke on-task attention mid-flow and sought praise for his efforts before continuing.

U2's breaks in attention are highlighted in his construction times which show only very small improvements over sessions. Specifically a mean drop of 3.2 seconds between session 1 and session 2, and a further 1.48 seconds between session 2 and 3. This contrasts markedly with say U1 who achieved a mean drop of 41.56 seconds between sessions 1 and 2. Indeed U2 made no significant improvement in construction speed between sessions. U2 did however manage to better the sample norms on his third session by 6.64 seconds. Although slightly bettering the sample norms this improvement was not significant.

In terms of efficiency U2's performance was slightly better, despite again using the same potentially inefficient 'key tapping' cursor movement strategy as U1. In contrast to U1, U2 did occasionally overshoot his target when moving the cursor too rapidly, and again one may be attributable to his short lapses in attention. Despite these momentary lapses he made significant gains in efficiency between session 2 and session 3. U2's efficiency was also significantly better than the sample norms by the completion of his third session.

Although U2 made significant improvements between some sessions and sample norms, he did not significantly improve over the three sessions when considered as a whole. This is somewhat surprising since efficiency measures are unaffected by elapsed time taken in task completion. In U2's case it is plausible, given his background, to suggest that lapses in attention not only slowed construction time but also had a knock on effect in terms of efficiency. Perhaps following a lapse, U2 found it difficult to re-focus on the task and in turn had to re-plan his navigational strategy, correcting for any errors made during the lapse. Thus overall, efficiency may have been reduced.

Given U2's poor performance, the parallel to his limited attention span must be drawn. Indeed one can attribute much of his poor performance improvement to his lapses in on-task attention. Observers who were present from time-to-time, commented on how his attentional lapses detracted from his apparent good progress with Easy Speaker. The comment was also made that they thought that his lapses were slightly more frequent than normal, but were not unexpected given the situation. It is hoped that as U2 uses Easy Speaker more frequently his on-task attention will improve. This will be examined more closely in the following chapter when all four users are rated by observers before and after Easy Speaker intervention using an amalgam of items from the Adaptive Behaviour Scales and Function Performance Record questionnaires.

Despite U2's performance he had a good understanding for the layout of the hierarchy and could along with U1 describe the navigational strategies he was using. This suggested that he had learnt sizeable chunks of the hierarchy and employed various planning methodologies in navigation.

USER 3: KC

Fig 9.33, Descriptive statistics for U3's ESAT sessions for U3

Task No.	Transcription of utterence to construct and generate	Session 1 TOT T	Session 1 TMC	Session 1 WMIN	Session 2 TOT T	Session 2 TMC	Session 2 WMIN	Session 3 TOT T	Session 3 TMC	Session 3 WMIN
1	meals	75.30	145.00	0.80	58.71	211.00	1.02	27.41	92.00	2.19
2	summer	146.65	445.00	0.41	135.12	500.00	0.44	30.54	107.00	1.96
3	trainers	185.32	428.00	0.32	130.07	531.00	0.46	62.45	137.00	0.96
4	muscles	268.20	749.00	0.22	113.04	269.00	0.53	91.50	220.00	0.66
5	ham/salad	194.15	541.00	0.62	112.16	268.00	1.07	66.07	249.00	1.82
6	read/library/book	217.34	719.00	0.83	246.83	809.00	0.73	92.00	301.00	1.96
7	is it/raining	193.89	466.00	0.93	249.42	1350.00	0.72	108.26	386.00	1.66
8	get the nurse/sore/back	272.37	835.00	1.10	208.06	1021.00	1.44	88.92	438.00	3.37
	Mean	194.15	541.00	0.65	156.68	619.88	0.80	70.89	241.25	1.82
	SD	63.92	222.26	0.31	69.72	408.15	0.35	29.77	128.12	0.82

Total time (TOT T)

Fig 9.34, Mean time (TOT T) taken to construct an ESAT task by session and norms for U3

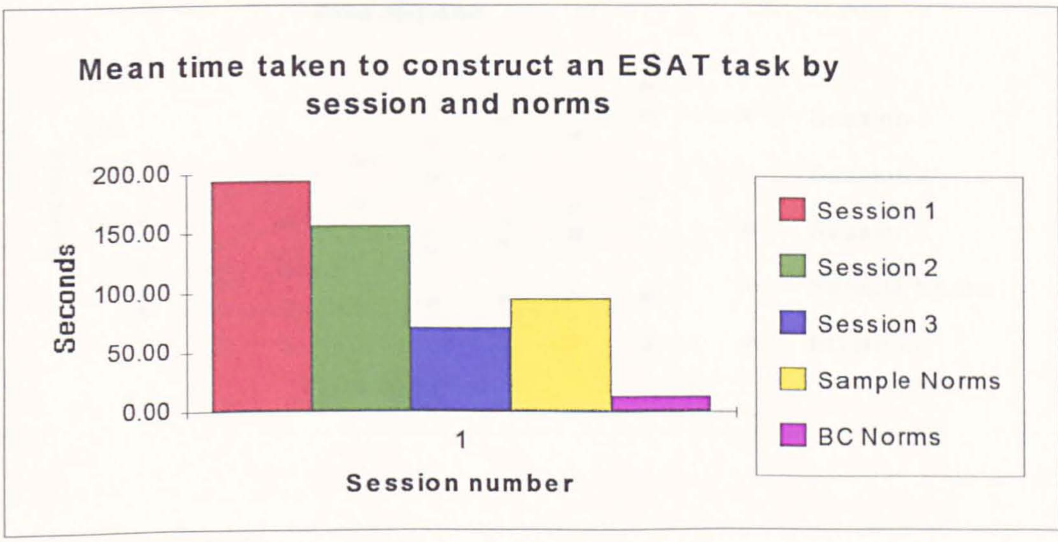


Fig 9.35, Wilcoxon Z matrix for total time (TOT T) for U3

	Session 1	Session 2	Session 3	S Norms	BC Norms
Session 1	-	-	-	-	-
Session 2	**-1.40**	-	-	-	-
Session 3	**-2.52****	**-2.52****	-	-	-
S Norms	-2.52**	-2.52**	**-2.52****	-	-
BC Norms	-2.52**	-2.52**	**-2.52****	-2.52**	-

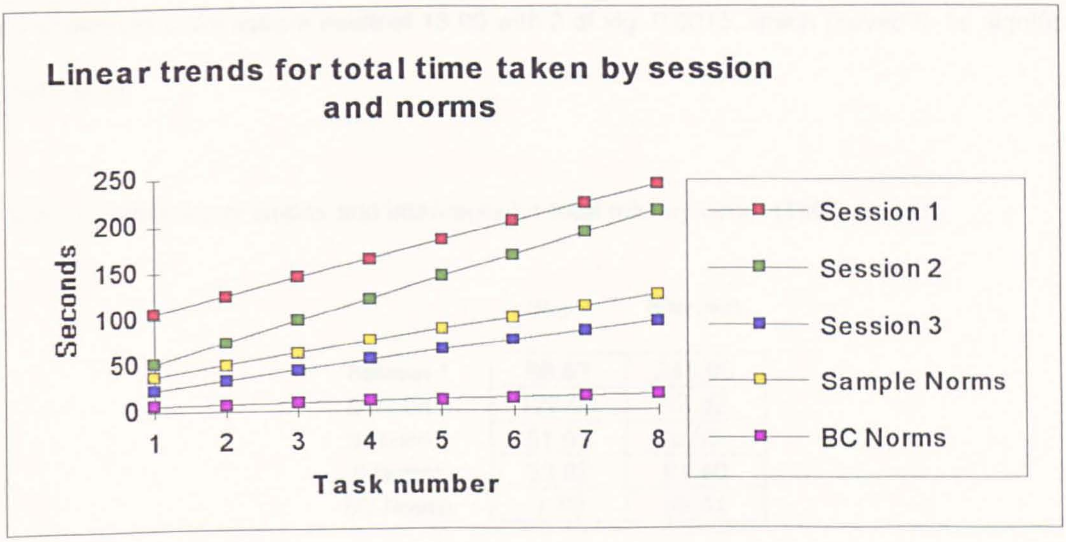
Key to significance levels: * 0.05, ** 0.01

To confirm there was a significant improvement over the three sessions a Friedman test was carried out yielding a Chi square value of 13.00 with 2 df sig. 0.0015, which proved to be significant at the .001 level.

Fig 9.36, A summary of slopes and intercepts for total time (TOT T) for U3

	Slope	Intercept
Session 1	19.50	106.42
Session 2	23.41	51.34
Session 3	10.50	23.62
S Norms	12.50	38.28
BC Norms	1.42	6.71

Fig 9.37, Calculated total time (TOT T) linear trends for each session and norms for U3



Total mickey count (TMC)

Fig 9.38, Mean total mickey count (TMC) taken to construct an ESAT task by session and norms for U3

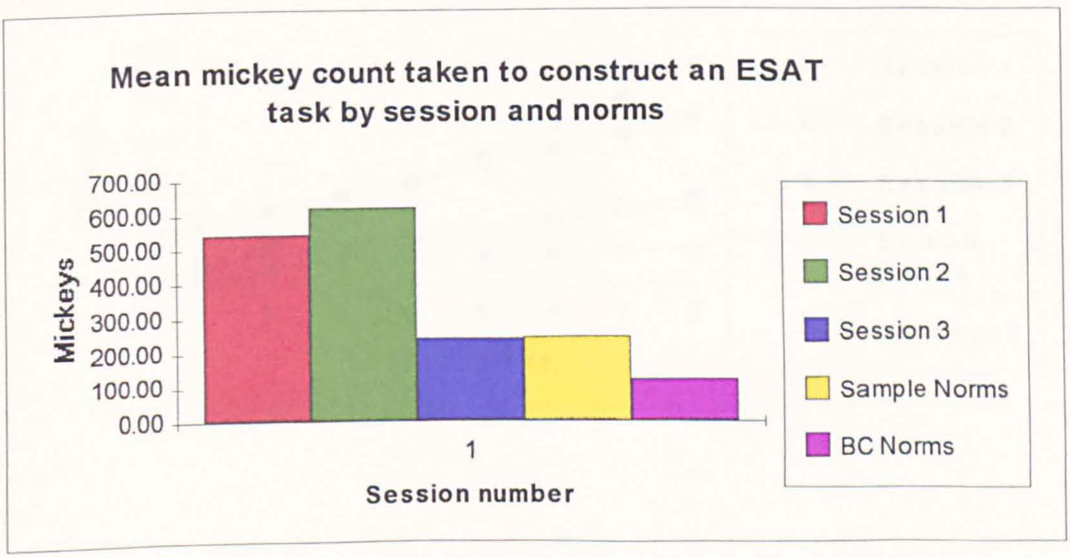


Fig 9.39, Wilcoxon Z matrix for total mickey count (TMC) for U3

	Session 1	Session 2	Session 3	S Norms	BC Norms
Session 1	-	-	-	-	-
Session 2	**-0.70**	-	-	-	-
Session 3	**-2.52****	**-2.52****	-	-	-
S Norms	-2.52**	-2.52**	**-0.42**	-	-
BC Norms	-2.52**	-2.52**	**-2.24***	-2.52**	-

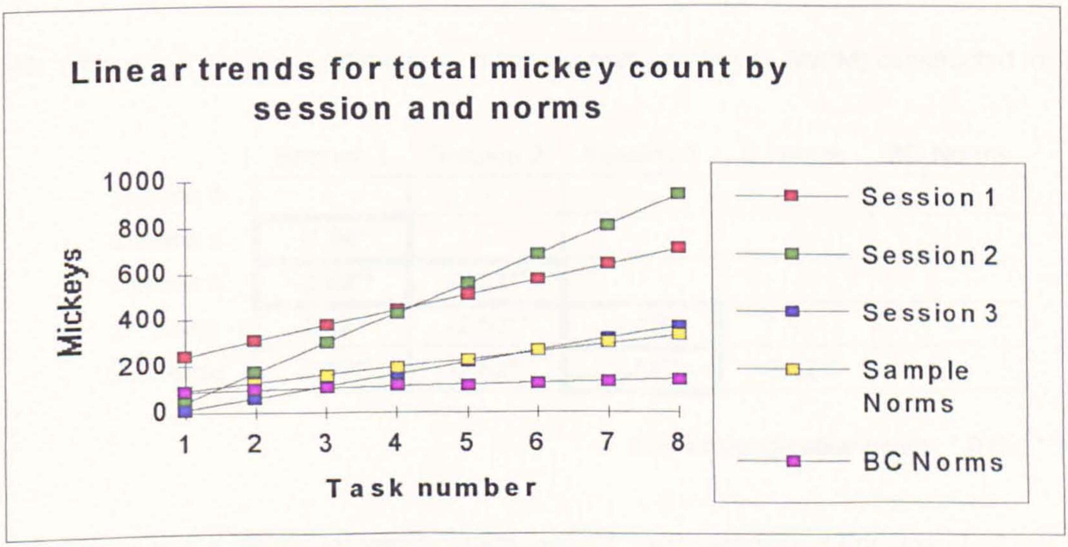
Key to significance levels: * 0.05, ** 0.01

To confirm there was a significant improvement over the three sessions a Friedman test was carried out yielding a Chi square value of 13.00 with 2 df sig. 0.0015, which proved to be significant at the .001 level.

Fig 9.40, A summary of slopes and intercepts for total mickey count (TMC) for U3

	Slope	Intercept
Session 1	66.67	241.00
Session 2	128.01	43.82
Session 3	51.64	8.86
S Norms	33.62	94.40
BC Norms	7.89	88.41

Fig 9.41, Calculated total mickey count (TMC) linear trends for each session and norms



Words per minute (WPM)

Fig 9.42, Mean number of estimated words per minute (WPM) constructed by session and norms for U3

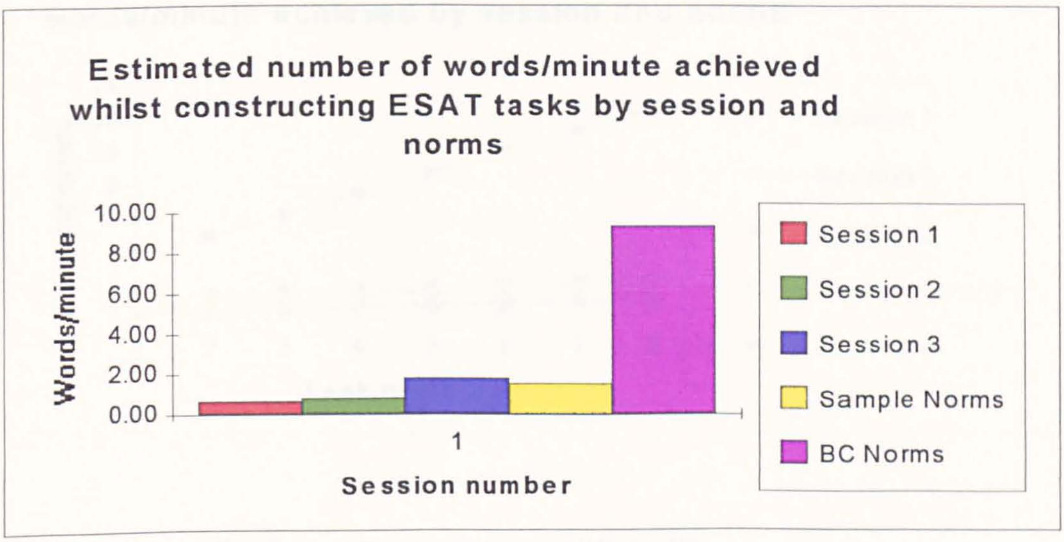


Fig 9.43, Wilcoxon Z matrix for estimated number of words per minute (WPM) constructed for U3

	Session 1	Session 2	Session 3	S Norms	BC Norms
Session 1	-	-	-	-	-
Session 2	**-1.96***	-	-	-	-
Session 3	**-2.52****	**-2.52****	-	-	-
S Norms	-2.52**	-2.52**	**-2.10***	-	-
BC Norms	-2.52**	-2.52**	**-2.52****	-2.52**	-

Key to significance levels: * 0.05, ** 0.01

To confirm there was a significant improvement over the three sessions a Friedman test was carried out yielding a Chi square value of 13.00 with 2 df sig. 0.0015, which proved to be significant at the .001 level.

Fig 9.44, A summary of slopes and intercepts for estimated number of words per minute constructed (WPM) for U3

	Slope	Intercept
Session 1	0.08	0.30
Session 2	0.08	0.49
Session 3	0.13	1.24
S Norms	0.13	0.68
BC Norms	1.26	3.65

Fig 9.45, Estimated number of words per minute (WPM) constructed linear trends for each session and norms for U3

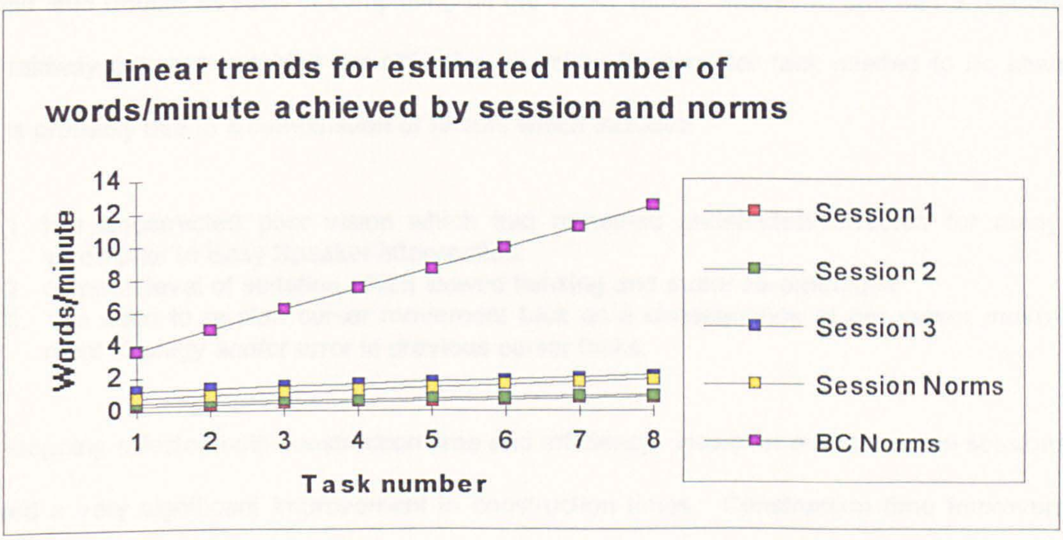


Fig 9.46, Summary of ESAT performance improvements for U3

Differences in ...	Significance Level			Significant Improvement		Improvement	
	NS	0.05	0.01	Yes	No	Yes	No
Speed between session 1 and 2	✓				✓	✓	
Speed between session 1 and 3			✓	✓		✓	
Speed between session 2 and 3			✓	✓		✓	
Speed between session 3 and sample norms			✓	✓		✓	
Speed between session 3 and best case norms			✓		✓		✓
Cursor tack efficiency between session 1 and session 2	✓				✓		✓
Cursor tack efficiency between session 1 and 3			✓	✓		✓	
Cursor tack efficiency between session 2 and 3			✓	✓		✓	
Cursor tack efficiency between session 3 and sample norms	✓				✓	✓	
Cursor tack efficiency between session 3 and best case norms		✓			✓		✓
Estimated no. of words/minute between session 1 and session 2		✓		✓		✓	
Estimated no. of words/minute between session 1 and 3			✓	✓		✓	
Estimated no. of words/minute between session 2 and 3			✓	✓		✓	
Estimated no. of words/minute between session 3 and sample norms		✓		✓		✓	
Estimated no. of words/minute between session 3 and best case norms			✓		✓		✓

Discussion for User 3: KC

U3 had little trouble as such in completing all the ESAT tasks. However, she had a tendency to stop midway between reaching her objective or point where cursor tack needed to be changed. This is probably due to a combination of factors which included :-

1. Her uncorrected poor vision which had remained undetected/corrected for many years prior to Easy Speaker intervention.
2. General level of sedation which slowed thinking and motor co-ordination.
3. The need to re-plan cursor movement tack as a consequence of her cursor movement strategy and/or error in previous cursor tacks.

This stopping effected both construction time and efficiency. However over the three sessions she showed a very significant improvement in construction times. Construction time improvements are illustrated when one examines mean drops between sessions. Specifically a mean drop of 37.47 seconds between session 1 and 2 and a further drop of 85.79 seconds between 2 and 3. By session 3 construction time were significantly better than sample norms. It is interesting to note that U3 continued to make large improvements in construction time right up to session 3. This suggests that unlike other users U3 would be likely to have continue this trend for improvement potentially over several more sessions.

As regards efficiency, as distinct from U1 and U2, U3 used a potentially faster and more efficient cursor movement strategy. This involved preplanning cursor tack goals, keeping the required cursor key depressed and then when the required goal was in sight releasing the key and switching to a more finely controlled key tapping method. However in U3's case this method led to a set of inconsistent cursor tack results, although not effecting construction times markedly. Specifically a mean increase of 105.00 mickeys between session 1 and session 2, but a drop of 378.68 mickeys between session 2 and 3. To help account for this uncharacteristic result it should be pointed out that the auto-repeat feature of the keyboard can move the cursor at high speed and did frequently cause U3 to overshoot the target. This occurred less on session 1 but by session 2 U3's overconfidence may have led her to attempt to squeeze the last ounce of performance from this technique and thus be prone to overshoot the target. Alternatively it could have been a result of her generally depressed state of thinking and motor co-ordination. However this second assumption is brought into question as she could readily correct these overshoots which suggest little depression of cognitive skills. Indeed her construction time was imperceptibly affected by these frequent errors. By session 3 it is felt that U3 had learned to anticipate where to switch strategies and avoid costly efficiency errors.

Given the rapid rates of improvement shown and the fact U3 could describe parts of the hierarchy visited and how to get there verbally, this suggested a good degree of learning. She was also aware of her errors in cursor tack and continually made efforts not to repeat them in later sessions.

USER 4: CP

Fig 9.47, Descriptive statistics for U4's ESAT sessions

Task No.	Transcription of utterence to construct and generate	Session 1			Session 2			Session 3		
		TOT T	TMC	WPM	TOT T	TMC	WPM	TOT T	TMC	WPM
1	meals	33.61	129.00	1.79	27.46	93.00	2.18	21.75	104.00	2.76
2	summer	46.19	131.00	1.30	43.83	152.00	1.37	33.17	91.00	1.81
3	trainers	69.15	143.00	0.87	49.49	127.00	1.21	40.31	125.00	1.49
4	muscles	98.10	227.00	0.61	84.53	154.00	0.71	54.05	139.00	1.11
5	ham/salad	70.03	236.00	1.71	51.20	234.00	2.34	46.02	166.00	2.61
6	read/library/book	108.48	163.00	1.66	85.96	259.00	2.09	51.85	136.00	3.47
7	is it/raining	97.93	252.00	1.84	76.85	288.00	2.34	56.46	221.00	3.19
8	get the nurse/sore/back	114.14	250.00	2.63	75.03	162.00	4.00	60.42	197.00	4.97
	Mean	79.70	191.38	1.55	61.79	183.63	2.03	45.50	147.38	2.67
	SD	29.61	54.84	0.63	21.61	68.53	1.00	13.09	44.71	1.24

Total time (TOT T)

Fig 9.48, Mean total time (TOT T) taken to construct an ESAT task by session and norms for U4

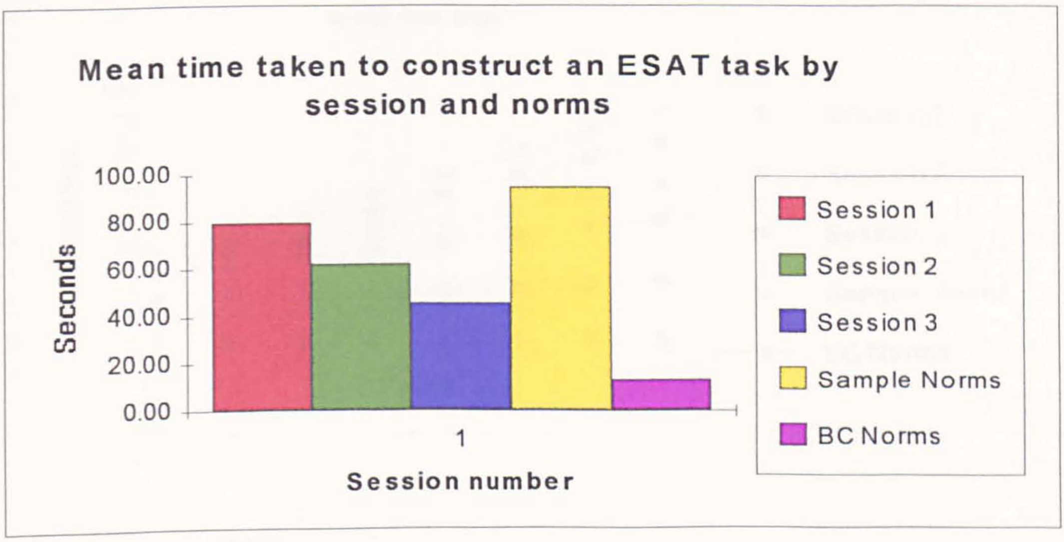


Fig 9.49, Wilcoxon Z matrix for total time (TOT T)

	Session 1	Session 2	Session 3	S Norms	BC Norms
Session 1	-	-	-	-	-
Session 2	**-2.52****	-	-	-	-
Session 3	**-2.52****	**-2.52***	-	-	-
S Norms	-2.52**	-2.52**	**-2.52****	-	-
BC Norms	-2.52**	-2.52**	**-2.52****	-2.52**	-

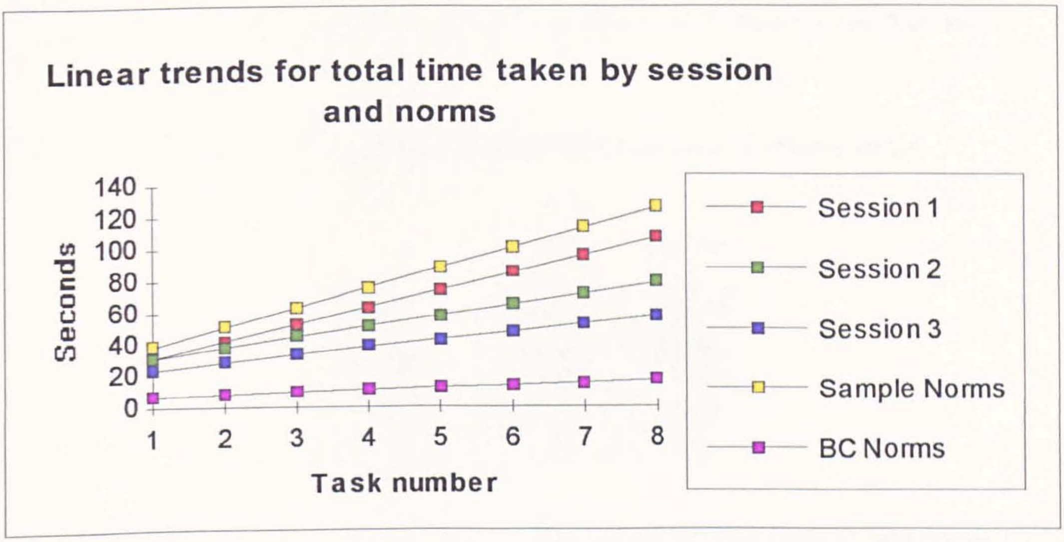
Key to significance levels: * 0.05, ** 0.01

To confirm there was a significant improvement over the three sessions a Friedman test was carried out yielding a Chi square value of 16.00 with 2 df which is significant at the .0005 level.

Fig 9.50, A summary of slopes and intercepts for total time (TOT T) for U4

	Slope	Intercept
Session 1	10.86	30.83
Session 2	6.84	31.03
Session 3	4.93	23.34
S Norms	12.50	38.28
BC Norms	1.42	6.71

Fig 9.51, Calculated linear trends for each session and norms for U4



Total mickey count (TMC)

Fig 9.52, Mean total mickey count (TMC) taken to construct an ESAT task by session an norms for U4

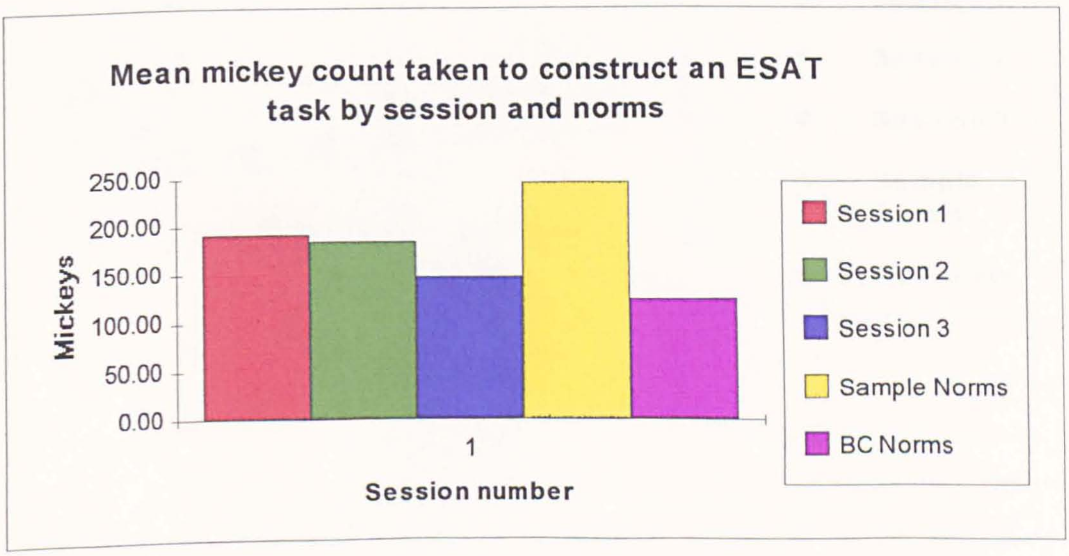


Fig 9.53, Wilcoxon Z matrix for total mickey count (TMC) for U4

	Session 1	Session 2	Session 3	S Norms	BC Norms
Session 1	-	-	-	-	-
Session 2	**-0.35**	-	-	-	-
Session 3	**-2.52****	**-1.68**	-	-	-
S Norms	-2.24*	-2.52**	**-2.52****	-	-
BC Norms	-2.52**	-2.10*	**-1.96***	-2.52**	-

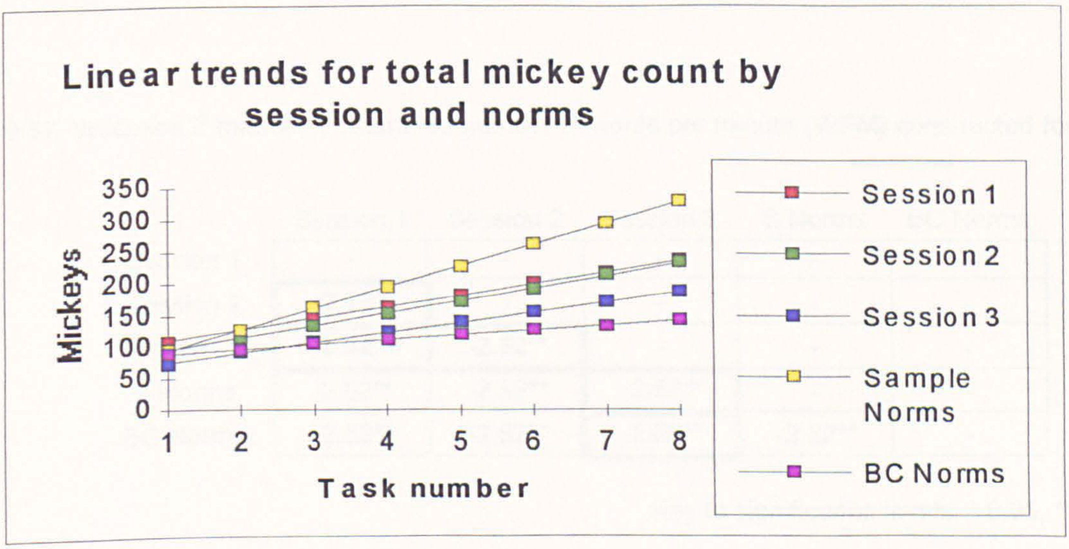
Key to significance levels: * 0.05, ** 0.01

To confirm there was a significant improvement over the three sessions a Friedman test was carried out yielding a Chi square value of 7.75 with 2 df which is significant at the 0.05 level.

Fig 9.54, A summary of slopes and intercepts for total mickey count (TMC) for U4

	Slope (b)	Intercept (a)
Session 1	18.11	109.89
Session 2	19.51	95.82
Session 3	16.20	74.46
S Norms	33.62	94.40
BC Norms	7.89	88.41

Fig 9.55, Calculated total mickey count (TMC) linear trends for each session and norms for U4



Words per minute (WPM)

Fig 9.56, Mean number of words per minute (WPM) constructed by session and norms for U4

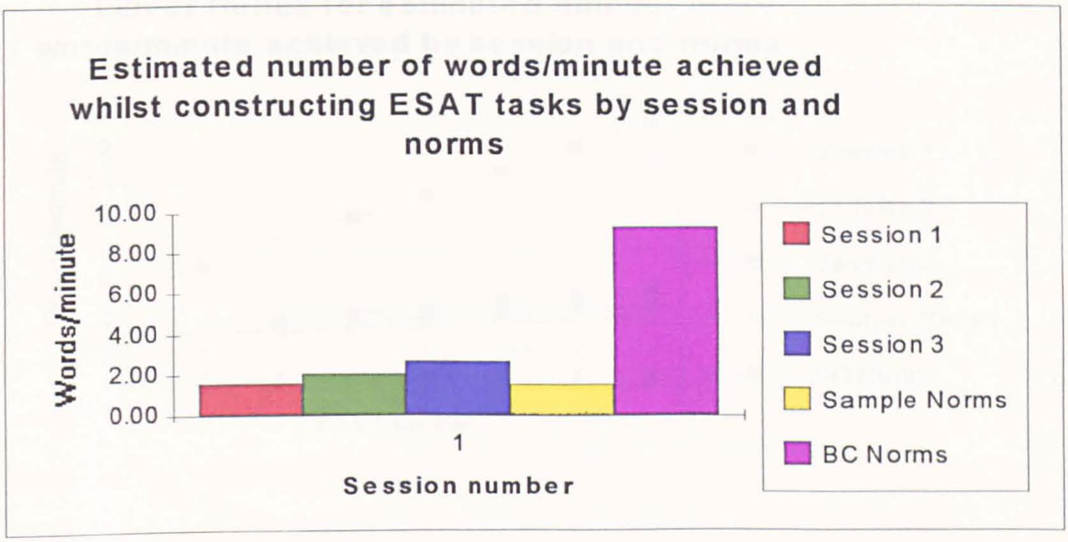


Fig 9.57, Wilcoxon Z matrix for estimated number of words per minute (WPM) constructed for U4

	Session 1	Session 2	Session 3	S Norms	BC Norms
Session 1	-	-	-	-	-
Session 2	**-2.52****	-	-	-	-
Session 3	**-2.52****	**-2.52****	-	-	-
S Norms	-2.52**	-2.52**	**-2.52****	-	-
BC Norms	-2.52**	-2.52**	**-2.52****	-2.52**	-

Key to significance levels: * 0.05, ** 0.01

To confirm there was a significant improvement over the three sessions a Friedman test was carried out yielding a Chi square value of 16.00 with 2 df sig. which is significant at the 0.0005 level.

Fig 9.58, A summary of slopes and intercepts for estimated number of words per minute constructed (WPM) for U4

	Slope	Intercept
Session 1	0.14	0.90
Session 2	0.26	0.86
Session 3	0.35	1.08
S Norms	0.13	0.68
BC Norms	1.26	3.65

Fig 9.59, Estimated number of words per minute (WPM) constructed linear trends for each session and norms for U4

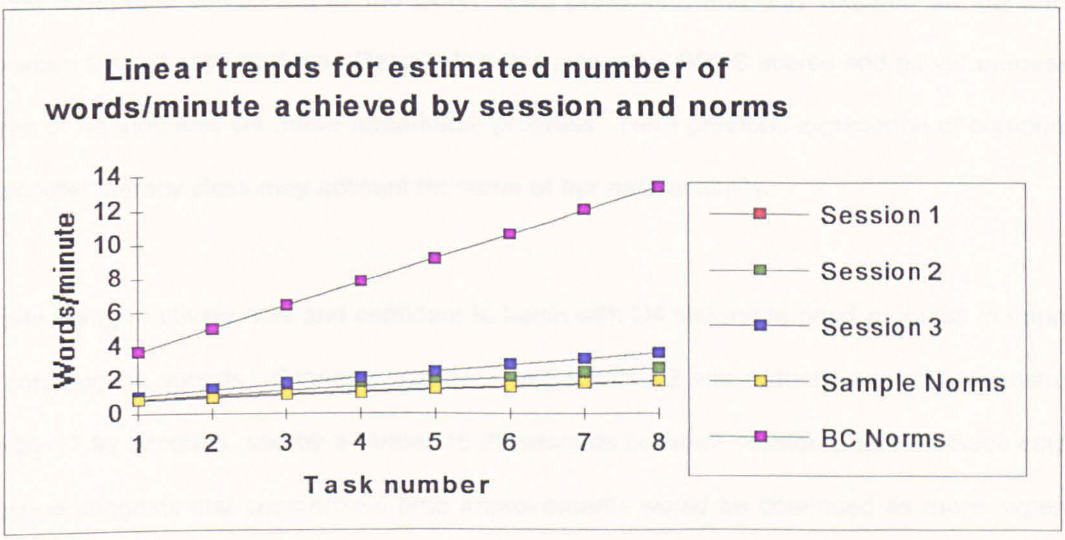


Fig 9.60, Summary of ESAT performance improvements for U4

Differences in ...	Significance Level			Significant Improvement		Improvement	
	NS	0.05	0.01	Yes	No	Yes	No
Speed between session 1 and 2			✓	✓		✓	
Speed between session 1 and 3			✓	✓		✓	
Speed between session 2 and 3		✓		✓		✓	
Speed between session 3 and sample norms			✓	✓		✓	
Speed between session 3 and best case norms			✓		✓		✓
Cursor tack efficiency between session 1 and session 2	✓				✓	✓	
Cursor tack efficiency between session 1 and 3			✓	✓		✓	
Cursor tack efficiency between session 2 and 3	✓				✓	✓	
Cursor tack efficiency between session 3 and sample norms			✓	✓		✓	
Cursor tack efficiency between session 3 and best case norms		✓			✓		✓
Estimated no. of words/minute between session 1 and session 2			✓	✓		✓	
Estimated no. of words/minute between session 1 and 3			✓	✓		✓	
Estimated no. of words/minute between session 2 and 3			✓	✓		✓	
Estimated no. of words/minute between session 3 and sample norms			✓	✓		✓	
Estimated no. of words/minute between session 3 and best case norms			✓		✓		✓

DISCUSSION FOR USER 4: CP

U4 had no trouble completing all the ESAT tasks presented, and only required the minimum of help when first attempting them. Despite her relatively poor BPVS scores and as yet unmeasured degree of hearing loss U4 made remarkable progress. Here previous experience of computers in a computer literacy class may account for some of her *natural* ability.

Despite being relatively able and confident to begin with U4 still made good progress in improving her construction speeds. Between session 1 and session 2 she reduced her mean construction time by 17.91 seconds, and by a further 16.29 seconds between session 2 and 3. Such continual reduction suggests that construction time improvements would be continued as more experience was gained. Given the high construction speeds achieved by U4 it is also likely that she might also approach the construction times achieved by an unimpaired individual. This assumption is given support by the very significant improvements she made between sessions.

U4 used a similar cursor movement strategy to U3 but this time to greater effect. Once a cursor tack was planned she would utilise the auto-repeat feature of the keyboard and move at high speed toward the target, taking very little time to change tack when the need arose. Her efficiency scores however are not as high as U1 who by session 3 was on par with an unimpaired individual. This is due to the odd slip which took the cursor too far off tack. This only occurred once or twice per session but was enough to reduce overall efficiency. Despite a slightly reduced efficiency level she made significant improvements over the three sessions. Had it not been for those occasional lapses she would undoubtedly reached the efficiency levels of unimpaired individuals. Support for this assertion is given from the fact that she appeared, despite her handicaps, to be using the same movement and planning strategies as unimpaired individuals, and with virtually the same effectiveness.

General Results

Despite users each having their own unique disabilities and collectively sharing a mean age equivalency of 5 years, their ability to use Easy Speaker is remarkable. One must remember that one cannot simply compare the ability of an unimpaired 5 year old individual with the ability shown by those in the sample. Each made definite performance gains within their own ability, and three of the four made statistically significant gains between sessions. Although evaluation has focused on performance improvements within individual users, it is still worthwhile comparing improvements between users. That is, not to statistically compare performance between users but rather to 'eyeball' the data. Obviously statistical comparisons between users are meaningless given their differing handicaps and abilities. One can appreciate these differing abilities when session means for each user are presented in table form :-

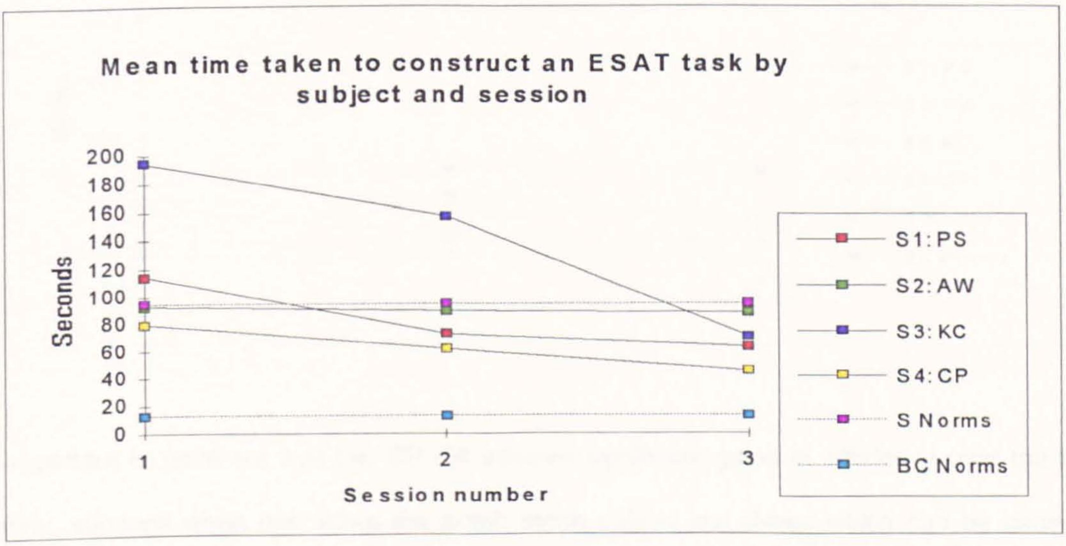
Fig 9.61, Summary of ESAT session means by user

	Session 1			Session 2			Session 3		
	TOT T	TMC	WPM	TOT T	TMC	WPM	TOT T	TMC	WPM
U1: PS	114.45	180.25	1.05	72.89	133.63	1.76	63.34	123.63	1.83
U2:AW	92.55	220.75	1.28	89.35	200.88	1.33	87.87	164.88	1.39
U3: KC	194.15	514.00	0.65	156.68	619.88	0.80	70.89	241.25	1.82
U4: CP	79.70	191.38	1.55	61.79	183.63	2.03	45.50	147.38	2.67
Mean	120.21	276.60	1.13	95.18	284.51	1.48	66.90	169.29	1.93
SD	51.34	159.19	0.38	42.54	225.39	0.54	17.57	50.87	0.54
S Norms	94.51	245.71	1.51	94.51	245.71	1.51	94.51	245.71	1.51
BC Norms	13.11	123.92	9.82	13.11	123.92	9.82	13.11	123.92	9.82

For example, one can see how U4: CP, was nearly 2.5 times faster in producing the ESAT tasks than user U3: KC who was the slowest on session one. She was also over 2.5 times more efficient. However U4: CP achieved an age equivalency of 4 years on the BPVS whereas U3: KC achieved 6 years 7 months. This apparent disparity in BPVS scores highlights why predictions and comparisons between individuals are fraught with the danger of being led toward invalid assumptions. Logically, one might have predicted that the higher BPVS score would represent higher overall ability and in turn lead to better Easy Speaker performance. However, in this case it should be pointed out that U4: CP had a good working knowledge of computers and had been in a computer literacy class previously.

To more easily compare performance between users it is useful to plot mean construction times by session for each user :-

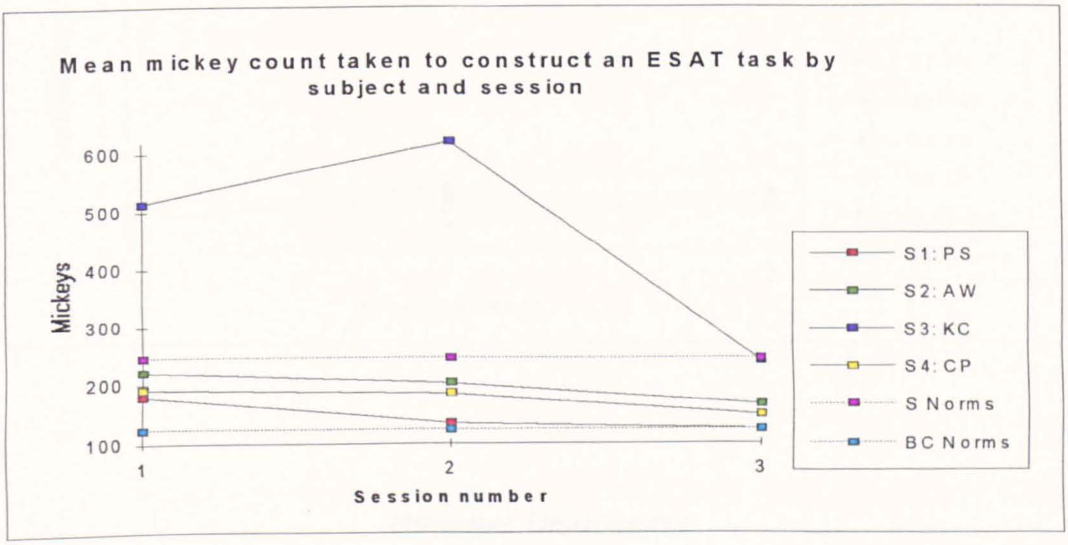
Fig 9.62, Mean time taken to construct an ESAT task by user and session



From the graph it is easy to see that one user (U2: AW) who failed to make significant improvements still made very small gains however; he didn't remain static or degrade in performance terms. Whereas U3: KC made very significant gains compared with her initial performance. It is important to remember is that any gains, significant or not are still worthwhile given the nature and abilities of the sample. In addition from the graph one can appreciate that each user potentially could have continued to make performance gains with the plot for U3: KC showing the most potential. However, plots of the other three users suggests a gradual 'bottoming out' of performance according to the law of diminishing returns as one might expect

In efficiency terms again three out of four users made significant efficiency gains. U2: AW again narrowly failed to make significant gains. However in efficiency terms his performance only narrowly failed to achieve significance. Indeed when individual sessions were evaluated he had made significant gains on three out of five session comparisons. Efficiency gains for the sample as a whole are highlighted when one views the following plot :-

Fig 9.63, Mean mickey count taken to construct an ESAT task by user and session

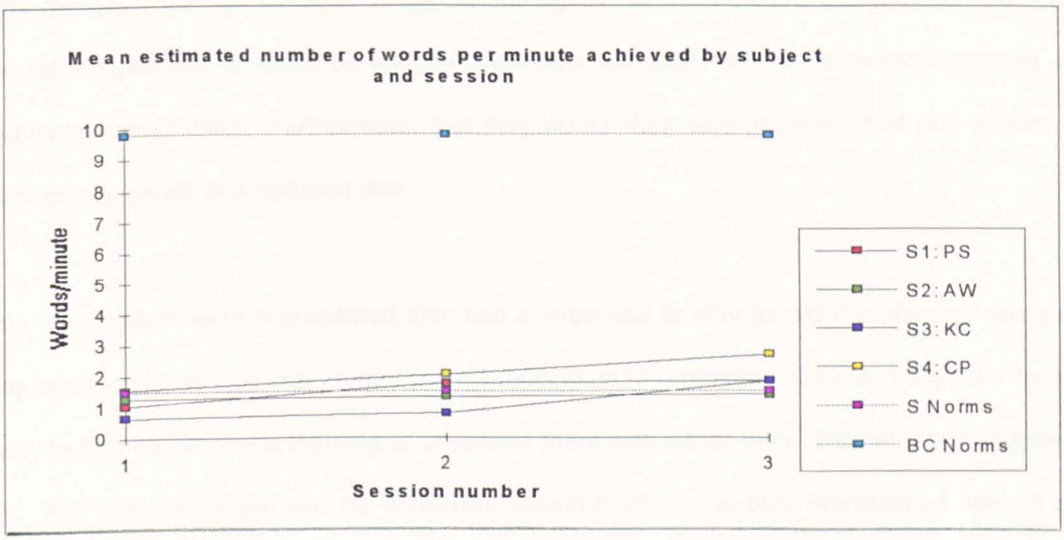


It is important to point out that U4: CP did achieve significant gains in efficiency over the three sessions, although when one views the graph mean values are shown which can be deceptive when viewed in isolation from the raw data.

U1: PS's efficiency by session three is equal to, if not slightly better, than that of an unimpaired individual on the same ESAT tasks. One must remember that efficiency is unrelated to speed on construction; in U1's case he was third fastest in session one and by session three was second fastest, as compared with other less efficient users.

Three of the four users also made significant gains in the number of words per minute (estimated) they produced. Again, U2: AW, narrowly failed to make significant gains despite his having made gains with each session. The generation rates are highlighted when one views the following graph :-

Fig 9.64, Mean estimated number of words/minute constructed by user and session



GENERAL DISCUSSION

As all that users acquired the skills needed to operate Easy Speaker quickly and efficiently one must logically assume that their capacity for skill acquisition remains intact despite their LD. This may or may not mean that other skills might be learnt to the same proficiency or with the same speed but rather the aptitude for doing so remains. This aptitude may only surface given optimum conditions. Conditions would be likely to encompass motivation, task, setting, physiology etc... Indeed such a notion is central to the 1992 AMMR reclassification of Learning Disability, and its optimistic outlook.

Given that all users could easily drive Easy Speaker to produce meaningful output this supports the notion that the hypermedia metaphor was appropriate. This metaphor proved to be so transparent that no users actually had any major problems with it. As suggested by the more complex ESAT tasks all users could successfully navigate between levels in the hierarchy, selecting symbols from each as they went ready to be output as digitised speech. Obviously the ability to expand the size of the hierarchy and increase interconnectivity is crucial for any dynamic screen approach from which multi-symbol utterances can be constructed.

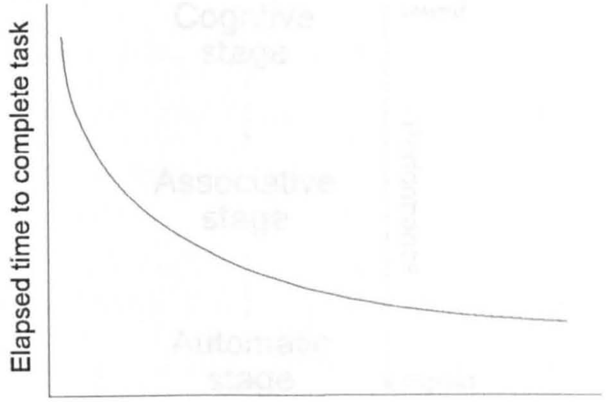
Despite this particular user group being significantly more able than those Easy Speaker was originally designed for, the strength of the psychological rationale behind the hypermedia metaphor would suggest that it would be equally applicable for those with more severe learning and communication disabilities. Furthermore, that they would show similar patterns of skill acquisition and productivity, albeit at a reduced rate.

The way the symbols were represented also had a large role to play as did the physical design of the user interface itself. All users found the symbols used to represents links and digitised speech 'obvious', or had no problems learning to associate them with either their destination or underlying speech. For example, there was no confusion between which symbols represented speech and which represented links to other parts of the hierarchy. It is thought that the high quality, concrete, images played a large role in the lack of confusion between the meaning of various symbols.

From the performance of users one can also propose that the symbol density was appropriate otherwise efficiency scores would have been likely to remain static due to scanning. That is, due to limited memory spans users would need to scan through symbols sequentially, or randomly, in order to locate the target symbol thus reducing efficiency. However, it is proposed that the layout of the screen facilitated memory, and hence contributed to increasing efficiency over sessions. A higher symbol density would have been less likely to produce a similar result due to a higher memory and cognitive load.

It terms of skill acquisition users, despite their obvious Learning Disabilities, displayed typical learning and efficiency curves. For example, a typical performance curve for learning to complete one of the ESAT tasks may resemble the one shown below :-

Fig 9.65, A hypothetical efficiency curve for elapsed time (TOT T) to complete an ESAT task

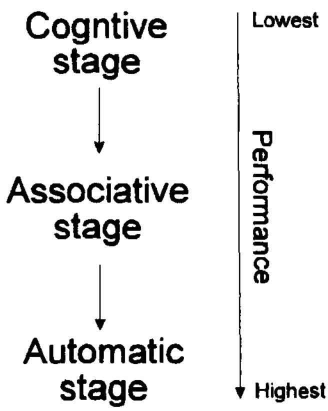


Elapsed time to complete a given task decreases as time spent honing that skill increases. Simply, the more time you devote to learning a skill the better, or faster, you will become. The one caveat being that you will never achieve perfection regardless of the time and effort devoted to learning. When one examines the efficiency curves produced by all users they fit this prototype well. Where improvements are marked early on and gradually decrease regardless of effort expended. Such a 'diminishing returns' curve suggests that users learnt to operate Easy Speaker relatively rapidly and then would be likely to hone their skill over time. Perhaps the 80:20 rule will illustrate the problems with performance improvements over prolonged periods. Simply according to the rule, 80 percent of the potential optimum performance will be achieved in 20 percent of the time devoted, and the last 20 percent of performance up to optimum will require 80 percent of the effort. Indeed optimum performance will invariably never be achieved.

As users display these characteristic curves we might reasonably assume that they are employing some of the same cognitive models and strategies unimpaired individuals might.

In the light of the pattern of skill acquisition shown one can postulate that users may have acquired their skills by virtue of the classic three stage process originally proposed by Fitts and Posner (1967). Although elaborated by later authors, e.g. Anderson, the original model meshes well with users skill patterns :-

Fig 9.66, The three stage model of skill acquisition



Skill acquisition begins with the Cognitive stage in which the user learns the *facts* associated with the skill domain. In this case the nut-and-bolts of moving the cursor, selecting speech symbols, navigating the hierarchy via links, internalising the hierarchy etc... Anderson (1983) suggests that people internally rehearse these facts as they attempt the newly *learned* task. That is, they apply domain-general problem-solving procedures which they have constructed, e.g. :-

.
.
.
IF the goal is to achieve state X
    and M is a method of achieving X
THEN set as a subgoal to apply M
.
.
.

From Johnson (1992)

In the case of Easy Speaker the following might be true :-

.
.
.
IF I need to generate the word "sugar"
    and I can get to the word "sugar" by selecting the link FOOD
THEN set as a subgoal selecting the link FOOD
.
.
.

Obviously to complete a given task many of these procedures would be constructed, combined, rehearsed and applied.

During the Associative stage any errors in initial understanding are reduced and hopefully eliminated. For example in the case of Easy Speaker a user may discover a quicker path through the hierarchy in order to reach their target. This being the case the user would at this stage utilise the new path in favour of the old. The most crucial feature of this stage is that connections between the items of knowledge are strengthened.

The function of the Associative stage is to furnish a successful procedure for performing the skill. That is, the declarative knowledge or facts have been translated into a procedural form of knowledge representation. The person can now apply these specific procedures resulting in increased performance and lower error rates. New rules can be added when required, old ones can be removed and more importantly existing rules can be merged and modified to produce more powerful procedures. That is, the procedures themselves are refined and optimised.

The final Automatic stage, as its name suggests, is where procedures from the associative stage are fully automated. That is, they become 'second nature', and as a result faster in their execution. When engaged in the skill we no longer have to *think* how to do it, almost like 'riding a bike'. Procedures are continually honed to accommodate exceptions and new strategies that are encounted and developed.

Given that one can apply a standard model of skill acquisition to a Learning Disabled population is encouraging. It means that we can apply specific teaching methods for training users to control and use Easy Speaker. For example, by concentrating on the Cognitive Stage, carers and health care professionals can impart the cognitive knowledge needed to make the best use of Easy Speaker, e.g. showing users how best to navigate to a given screen. This is opposed to trial and error, or discovery learning (Papert) in which users may develop many unproductive strategies which may be difficult to correct at a later date. Such incorrect strategies might be considered more detrimental for the Learning Disabled user than an unimpaired individual.

As users utilise Easy Speaker more, it is hoped that automaticity will develop enabling both construction speed and efficiency to increase to their optimum for that user. In addition the greater the level of automaticity the lower the level of cognitive effort that needs to be devoted to Easy Speaker and the greater the potential benefit Easy Speaker use might bring.

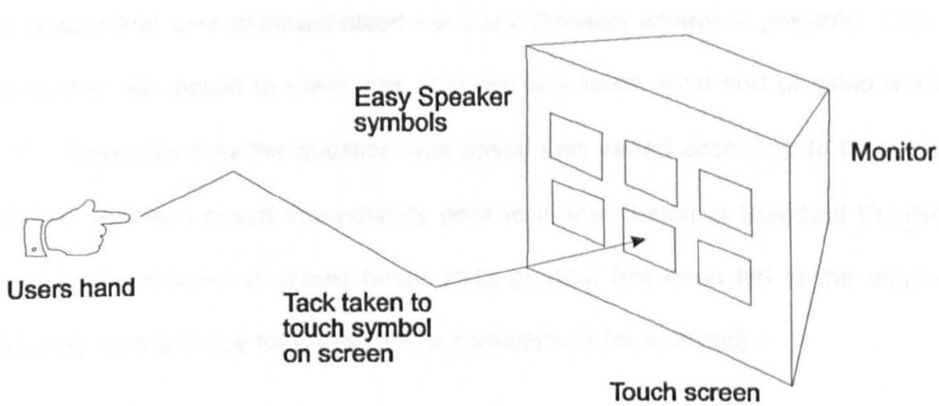
At this point is it salient to discuss the assertion made by Alm, Arnott and Newell (1992) that if a device is to be used as an ACD, users should be capable of producing more than the three words per minute which listeners find barely tolerable. Alm et al suggest a workable rate of 15 words per minute to allow message passing. Normal conversation has an average rate of 150-200 words per minute (Foulds 1980). Other studies have stressed how a negative attitude towards the speaker can build with silence (Newman 1982) and slow rate of communication (Scherer 1979).

With greater Easy Speaker skill automaticity one might reasonably expect construction speed and efficiency to improve, and to have the knock-on effect of increasing words per minute scores. Within this study, which was concerned only with skill acquisition, by session three users were able to produce an average of 1.93 words per minute, as compared with 9.82 wpm produced by an unimpaired individual. Obviously even the 9.82 wpm average of an unimpaired user is well below the 15 wpm suggested, and just a fraction of the 150-200 wpm of normal conversation. When viewed at face value all does not bode well for Easy Speaker being utilised as a augmentative communication device. However it should be stressed that 12 out of 14 speech symbols selected during ESAT tasks were single words, While of the remaining two, one contained two words ("is it"), and the other three ("get the nurse") . When viewed in this light it is not surprising that a low average wpm was achieved. Given the fact that Easy Speaker hierarchies can be constructed with speech symbols representing words/phrases with one to say 300 words a piece, using wpm scores for comparisons between communication systems can be misleading.

In a typical Easy Speaker hierarchy it would be likely that full phrases would be implemented under a single speech symbol, with just the keyword needing to be added, e.g. "For breakfast can I have ..." etc... In this case one might add "toast" making a six word utterance with just two speech symbols. One must remember though that if a symbol represents a phrase the symbol itself must be chosen very carefully so that it is representative. As a rough estimate one might suggest using this approach a 30-50 wpm output would be attainable by most users. Such speeds may seem high, but with the addition of a touch screen the process of symbol selection and navigation would be speeded up dramatically. Within the current study controlling Easy Speaker via the keyboard, as all users did, is inferior to symbol selection with a touch screen.

A touch screen was not used in the current study as it would have made measurement of efficiency gains, or cursor tack, near impossible. Obviously with a touch screen there would still be a tack taken by a users hand but measuring improvements in efficiency as compared with cursor tack would be highly unreliable.

Fig 9.67, Measuring hand movements, or tack, of users using a touch screen with Easy Speaker



Although the main purpose of both this chapter and the study discussed was to objectively measure performance improvements of users whilst learning to use Easy Speaker the output from ARTS during assessment was also analysed quasi-experimentally in a qualitative fashion (Coolican 1990, Elmes, Kantowitz and Roediger 1992). That is, during the study period all users ARTS output was combined so that a qualitative analysis of all the non-ESAT utterances produced could be made. This was carried out to check that users were using Easy Speaker to produce multi-symbol utterances that were semantically meaningful, if not grammatically correct, e.g. requests, statements and questions etc... If users were indeed generating purposeful utterances on-device this would suggest that Easy Speaker could be used as a primary means of communication by a non-vocal Learning Disabled population.

Qualitative analysis of output suggested that all users were using Easy Speaker generatively. As one might expect this was within the ranges of each users ability. For example, many utterances whilst semantically salient are ungrammatical, just as they would be if the user had uttered them using natural language. Obviously a symbolic system cannot impose or improve a grammatical rule base over what the user already posses within the limits of their ability.

From the actual ARTS data one can see how multi-symbol utterances were constructed. As mentioned earlier many of the sessions with Easy Speaker were clinically driven. That is, other than completion of set ESAT tasks, the output users produced was driven by them. Simply, the author may have directed a session so that particular parts of the hierarchy were visited, which may have lead to a branch elsewhere etc... One even led the next and so on ... However, the author attempted to ensure that user communicated via Easy Speaker wherever possible. One of the ad hoc questions that was posed to users was, "Do you like salad, what sort of salad would you like for dinner ?". Obviously how the question was posed was varied according to the abilities of the user in question and was posed immediately prior to dinner (lunch in Standard English). Actual ARTS output for three users is shown below (*NB printout has been left in the original Comma Separated Value format ready for import into a spreadsheet for analysis) :-

Fig 9.68a, U1's ARTS data for the "... what sort of salad would you like for dinner ? ..."

Utterance 2 :, ham/spring onion/lettuce/tomato/
Track :, [sys_start],<MAINMENU>,<FOODBLK>,<SALADS>,/HAM/,/SPRINGON/,/LETTUCE/,/TOMATO/,[output]
Speech Section Lengths :,0,0,0,0,1,2,1,1,0
Mickey Data :,0,17,1,31,25,13,16,13,1
Timings : ,11:37:48.60,11:38:06.18,11:38:48.14,11:39:18.18,11:39:40.43,11:39:55.48,11:40:06.19,11:40:20.74,11:40:50.84
Differences : ,0,17.58,41.96,30.04,22.25,15.05,10.71,14.55,30.10
Cumulative : ,0,17.58,59.54,89.58,111.83,126.88,137.59,152.14,182.24
Total Time :, 182.24
Total Mickey Count:, 117
Utterance Length :, 5
No of speech icons :, 4
No of link icons :, 3
No of Left Button Presses:, 7
No of Right Button Presses:, 1
Tot No of Button Presses:, 8
Ave Mickeys/Word :, 23.40
Ave Sec/Word :, 36.45
Ave Words/Min :, 1.65

Fig 9.68b, U2's ARTS data for the "... what sort of salad would you like for dinner ? ..."

Utterance 13 :, ham/salad/
Track :, [sys_start],<FOODBLK>,<SALADS>,/HAM/,/SALAD/,[output]
Speech Section Lengths :,0,0,0,1,1,0
Mickey Data :,0,71,51,34,32,1
Timings : ,13:21:53.13,13:22:03.08,13:22:31.69,13:22:54.93,13:23:13.44,13:23:15.96
Differences : ,0,9.95,28.61,23.24,18.51,2.52
Cumulative : ,0,9.95,38.56,61.80,80.31,82.83
Total Time :, 82.83
Total Mickey Count:, 189
Utterance Length :, 2
No of speech icons :, 2
No of link icons :, 2
No of Left Button Presses:, 4
No of Right Button Presses:, 1
Tot No of Button Presses:, 5
Ave Mickeys/Word :, 94.50
Ave Sec/Word :, 41.42
Ave Words/Min :, 1.45

Fig 9.68c, U4's ARTS data for the "... what sort of salad would you like for dinner ? ..."

Utterance 5 :, salad/tomato/lettuce/spring onion/
Track :, [sys_start],/SALAD/,/TOMATO/,/LETTUCE/,/SPRINGON/,[output]
Speech Section Lengths :,0,1,1,1,2,0
Mickey Data :,0,2,16,14,13,1
Timings : ,14:18:39.56,14:18:49.77,14:19:01.69,14:19:07.90,14:19:25.37,14:19:53.49
Differences : ,0,10.21,11.92,6.21,17.47,28.12
Cumulative : ,0,10.21,22.13,28.34,45.81,73.93
Total Time :, 73.93
Total Mickey Count:, 46
Utterance Length :, 5
No of speech icons :, 4
No of link icons :, 0
No of Left Button Presses:, 4
No of Right Button Presses:, 1
Tot No of Button Presses:, 5
Ave Mickeys/Word :, 9.20
Ave Sec/Word :, 14.79
Ave Words/Min :, 4.06

As one case see the actual answers to the question posed by the author varies according to the individual users ability. It is not simply a case of the user selecting all the relevant salad symbols shown on screen as some did not like certain salad items and so choose not to include them. U1 in this case left out the symbol salad as this was signified by a picture which predominantly showed lettuce which it was consequently found he didn't like. U4 choose not to include "Ham" as she considered that a salad should only consist of salad items and not meats etc... Other users such as U2, who test (BPVS) scores were lowest, simply constructed the utterance "Ham salad".

Further evidence that users could use Easy Speaker communicatively comes from the mid session breaks when users were asked what they would like to drink, after which both user and author went to fetch them. The following ARTS data was generated by users immediately prior to session breaks in response to the question of what they wanted to drink during the break ?

Fig 9.69a, U1's ARTS data for the "... what would you like to drink ? ..."

Utterance 1 :, tea/milk/sugar/sugar/
Track :, [sys_start],<FOODBLK>,<DRINKS>,/TEA2/,/MILK2/,/SUGAR/,/SUGAR/,[output]
Speech Section Lengths :,0,0,0,1,1,1,1,0
Mickey Data :,0,36,49,1,34,8,1,1
Timings : ,11:35:07.89,11:35:46.94,11:36:16.98,11:36:37.80,11:37:07.08,11:37:19.82,11:37:31.19,11:37:43.00
Differences : ,0,39.05,30.04,20.82,29.28,12.74,11.37,11.81
Cumulative : ,0,39.05,69.09,89.91,119.19,131.93,143.30,155.11
Total Time :, 155.11
Total Mickey Count:, 130
Utterance Length :, 4
No of speech icons :, 4
No of link icons :, 2
No of Left Button Presses:, 6
No of Right Button Presses:, 1
Tot No of Button Presses:, 7
Ave Mickeys/Word :, 32.50
Ave Sec/Word :, 38.78
Ave Words/Min :, 1.55

Fig 9.69b, U3's ARTS data for the "... what would you like to drink ? ..."

Utterance 2 :, tea/
Track :, [sys_start],<FOODBLK>,<DRINKS>,/TEA2/,[output]
Speech Section Lengths :,0,0,0,1,0
Mickey Data :,0,70,162,1,1
Timings : ,11:08:38.12,11:09:12.12,11:10:52.80,11:11:12.25,11:11:15.27
Differences : ,0,34.00,100.68,19.45,3.02
Cumulative : ,0,34.00,134.68,154.13,157.15
Total Time :, 157.15
Total Mickey Count:, 234
Utterance Length :, 1
No of speech icons :, 1
No of link icons :, 2
No of Left Button Presses:, 3
No of Right Button Presses:, 1
Tot No of Button Presses:, 4
Ave Mickeys/Word :, 234.00
Ave Sec/Word :, 157.15
Ave Words/Min :, 0.38

Fig 9.69c, U4's ARTS data for the "... what would you like to drink ? ..."

Utterance 1 :, tea/milk/
Track :, [sys_start],<FOODBLK>,<DRINKS>,/TEA2/,/MILK2/,[output]
Speech Section Lengths :,0,0,0,1,1,0
Mickey Data :,0,88,44,1,25,11
Timings : ,14:13:04.24,14:13:59.88,14:14:26.24,14:14:47.44,14:14:59.14,14:15:16.50
Differences : ,0,55.64,26.36,21.20,11.70,17.36
Cumulative : ,0,55.64,82.00,103.20,114.90,132.26
Total Time :, 132.26
Total Mickey Count:, 169
Utterance Length :, 2
No of speech icons :, 2
No of link icons :, 2
No of Left Button Presses:, 4
No of Right Button Presses:, 1
Tot No of Button Presses:, 5
Ave Mickeys/Word :, 84.50
Ave Sec/Word :, 66.13
Ave Words/Min :, 0.91

Utterance 2 :, sugar/
Track :, [sys_start],/SUGAR/,[output]
Speech Section Lengths :,0,1,0
Mickey Data :,0,44,1
Timings : ,14:15:19.63,14:15:35.78,14:15:47.42
Differences : ,0,16.15,11.64
Cumulative : ,0,16.15,27.79
Total Time :, 27.79
Total Mickey Count:, 45
Utterance Length :, 1
No of speech icons :, 1
No of link icons :, 0
No of Left Button Presses:, 1
No of Right Button Presses:, 1
Tot No of Button Presses:, 2
Ave Mickeys/Word :, 45.00
Ave Sec/Word :, 27.79
Ave Words/Min :, 2.16

Interestingly U2, who on test scores was the second most able, used Easy Speaker in a truly generative manner without prompting. He wanted to say "Tea with milk and two sugar's" but was constrained by the available symbol set and couldn't say "two sugar's" easily. In order to overcome this he simply selected the "sugar" symbol twice to indicate he wanted two sugars. Those in the canteen later confirmed that he did indeed usually have two sugars suggesting this had not been a chance generation.

U3 on the other hand who was most able based on test scores simply requested tea by selecting the "tea" symbol. This was typical of U4 who tended to "say" as little as possible in order to reduce effort on her part as she was the most physically disabled of the sample and had poor uncorrected vision. One can infer that what she actually meant was, "tea with one sugar", as this was what she normally had at break times and was known to the author who normally fetched it for her.

Again, suggesting that users were using Easy Speaker truly generatively is the ARTS data for U4 who when requesting "tea with milk and one sugar" forgot to include "sugar" within the original utterance before it was output as digitised speech. She resolved the situation some 30 seconds later with the output of the word "sugar" indicating that she had originally forgotten to include it, realised she needed to if she wanted sugar with her drink and selected the "sugar" symbol as a separate utterance.

Fig 9.70, The actual symbol screen from which users requested drinks, here shown with U3's request for tea

The two most capable users of Easy Speaker, namely U1 an U4 frequently demonstrated that they had a good grasp of system operation and could generate salient complex multi-symbol utterances which required them navigate between a number of different levels. U1 in response to rain beginning to fall on a skylight generated :-

Fig 9.71, Spontaneous generation of “is it raining ?” by U1

Utterance 19 :, is it/raining/
Track :, [sys_start],<GENERAL>,<WEATHER>,/ISIT/,<WEATHTYP>,/RAINING/,[output]
Speech Section Lengths :,0,0,0,2,0,1,0
Mickey Data :,0,94,42,14,36,31,1
Timings : ,13:29:04.79,13:29:19.57,13:29:43.96,13:29:58.95,13:30:20.92,13:31:12.77,13:31:14.97
Differences : ,0,14.78,24.39,14.99,21.97,51.85,2.20
Cumulative : ,0,14.78,39.17,54.16,76.13,127.98,130.18
Total Time :, 130.18
Total Mickey Count:, 218
Utterance Length :, 3
No of speech icons :, 2
No of link icons :, 3
No of Left Button Presses:, 5
No of Right Button Presses:, 1
Tot No of Button Presses:, 6
Ave Mickeys/Word :, 72.67
Ave Sec/Word :, 43.39
Ave Words/Min :, 1.38

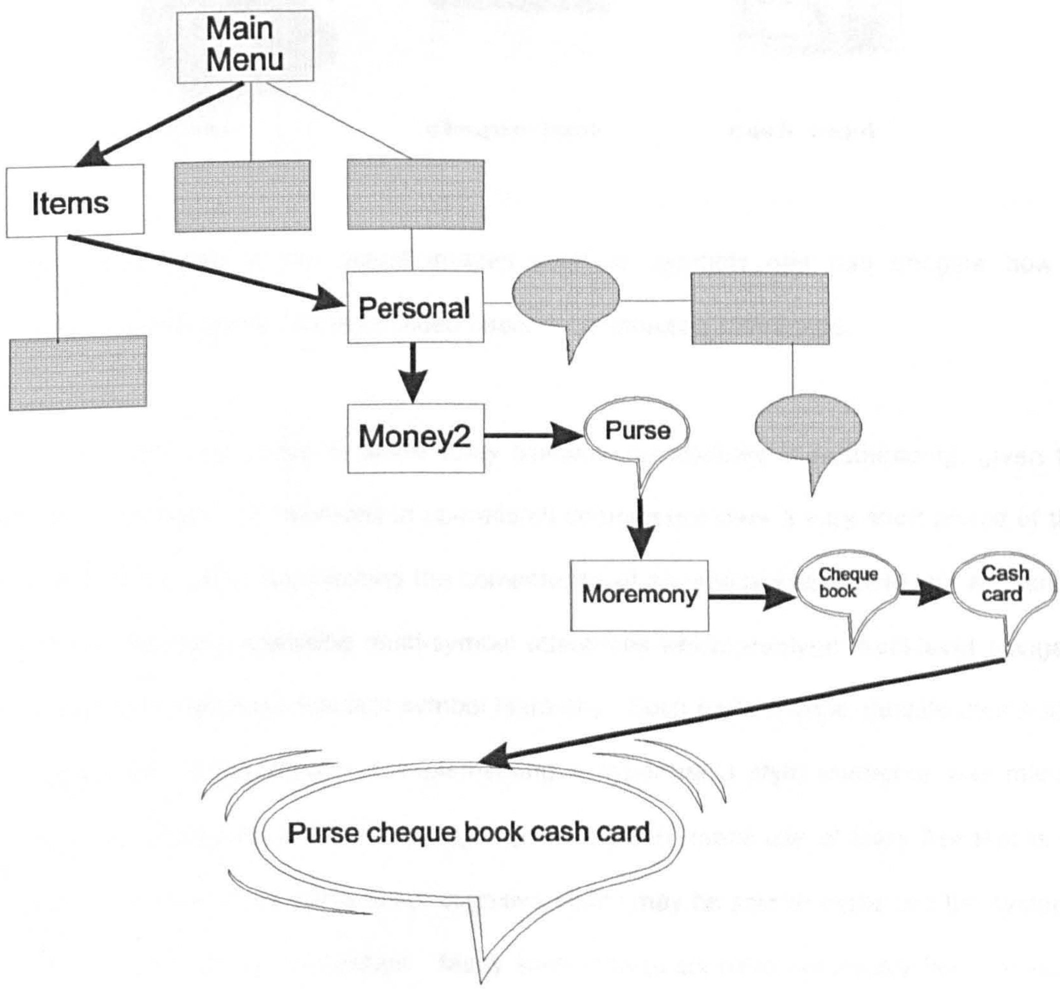
U4 who was the most behaviourally competent and independent functioning indicated that she was going shopping after the session and then went on to generate an utterance that had been “drilled” into her by various clinical psychologists and care assistants who see monetary awareness as being crucial to adaptive functioning :-

Fig 9.72, U4’s statement on what she needed to take shopping in order to buy things

Utterance 1 :, purse/cheque book/cash card/
Track :, [sys_start],<ITEMS>,<PERSONAL>,<MONEY2>,/PURSE/,<MOREMONY>,/CHECKBOO/,/CASHCARD/,[output]
Speech Section Lengths :,0,0,0,0,1,0,2,2,0
Mickey Data :,0,30,26,1,16,44,40,15,1
Timings : ,15:03:09.05,15:03:41.67,15:04:05.46,15:04:24.57,15:04:38.58,15:04:56.70,15:05:17.02,15:05:26.75,15:05:29.71
Differences : ,0,32.62,23.79,19.11,14.01,18.12,20.32,9.73,2.96
Cumulative : ,0,32.62,56.41,75.52,89.53,107.65,127.97,137.70,140.66
Total Time :, 140.66
Total Mickey Count:, 173
Utterance Length :, 5
No of speech icons :, 3
No of link icons :, 4
No of Left Button Presses:, 7
No of Right Button Presses:, 1
Tot No of Button Presses:, 8
Ave Mickeys/Word :, 34.60
Ave Sec/Word :, 28.13
Ave Words/Min :, 2.13

When one examines ARTS output more carefully one can see that users are not simply constructing multi-symbol utterances by combining symbols from one screen. They are instead navigating between screens in order to select symbols and construct meaningful utterances. U4 when constructing the utterance above had to visit no less than four different screens of symbols, selecting symbols from two of them. The actual path taken can be see more clearly in the concept map below :-

Fig 9.73, A concept map of the path U4 took in order to generate the utterance "purse cheque book cash card"



As indicated by the concept map above U4 had to navigate from the main menu to the "items" concept screen of symbols, then to "personal", then "money2", at which point she selected the speech symbol "purse", then navigated to "moremony", and finally selected the speech symbols "cheque book" and "cash card". The utterance was then output as digitised speech. In the diagram above some of the alternative concept screens and speech symbols she could have visited, alternative routes and other speech symbols are shaded in grey. Obviously the set of alternatives that was available was far greater than those shown on the schematic, which is intended to give some idea of the navigation steps involved in generating the utterance.

Fig 9.74, The utterance "purse cheque book cash card" as U4 assembled it immediately prior to output as digitised speech


purse


cheque-book


cash card

From the screen grab of the actual images used as symbols one can imagine how the "concreteness" of the symbol set itself aided users in constructing utterances.

Given that all users were able to utilise Easy Speaker successfully is encouraging, given their abilities and handicaps. All improved in operational competence over a very short period of time. Some even to the point of approaching the competence of an unimpaired individual. All were capable of spontaneously generating multi-symbol utterances which involved multi-level navigation and selection within the Easy Speaker symbol hierarchy. Such performance, despite their abilities would suggest that the approach of implementing a hypermedia style metaphor was relevant. Furthermore, the ease with which the user group managed to make use of Easy Speaker is suggestive that those non-vocal and of lower cognitive ability may be able to make use the system as their primary means of communication. Many such individuals have previously been precluded from ACD communication strategies due to their *perceived* limited ability and the complexity of many available devices.

SUMMARY

Given that three of the four users made significant performance gains over the three sessions one can readily accept that they acquired the skills necessary to achieve this. In the light of this skill acquisition it was proposed that learning disabled users might also acquire a skill using similar cognitive strategies to unimpaired users.

As *all* users *easily* acquired the necessary skills to use Easy Speaker this would suggest that in design terms the software is ideally matched to those of the samples ability; and furthermore that it's inherent motivating properties reinforced this skill acquisition. That is, in terms of the hyper-media style metaphor, the use of high quality photorealistic graphics, and use of CD quality digital audio and speech.

In addition to improving in pure performance terms all users also used Easy Speaker generatively as a Voice Output Communication Aid (VOCA). All were capable of spontaneously generating multi-symbol utterances which involved multi-level navigation and selection within the Easy Speaker symbol hierarchy.

The question that now needs to be answered is whether users who can learn to use Easy Speaker successfully can benefit in relation to off-device criteria. That is, does the acquisition of new linguistic and cognitive skills enable a corresponding improvement in everyday behavioural and communicative competence. Such changes can be assessed using adaptive behavioural techniques.

CHAPTER 10: Easy Speaker in use: Off-device improvements in language and cognition

Although users could acquire the skills necessary to operate and make productive use of a hypermedia-based Augmentative Communication Device, it was not demonstrably known whether any of the predicted off-device improvements in language and cognition would occur.

In order to investigate off-device gains an in-depth 24 week study of the four users outlined in chapter nine was conducted. Throughout this 24 week period users had weekly two hour sessions in which they utilised Easy Speaker.

To measure any gains over the 24 week period, immediately prior to the first session, a general Easy Speaker Progress Monitoring (ESPM) questionnaire was given to two of each users immediate caregivers to establish a baseline for key adaptive behaviours. This contained 69 relevant items taken from The Adaptive Behaviour Scales (ABS) and The Functional Performance Record 16 (FPR-16), and was subdivided into six domains.

After the final session a modified ESPM questionnaire was given to the same caregivers. In this questionnaire they were reminded of the responses they gave 24 weeks earlier and were asked to rate any improvement on each item. Any changes, either positive or negative, were then compared with the expected "no change" over baseline.

INTRODUCTION

Having addressed the question of whether users could acquire the necessary skills to operate and use Easy Speaker productively in chapter nine, it is to the question of whether the predicted off-device gains in communicative and other cognitive abilities actually occur with prolonged use in a borderline verbal group. Despite earlier single-subject studies suggesting Easy Speaker intervention might produce worthwhile gains in communicative and other cognitive abilities off-device (e.g. Rostron, Plant and Hermann 1994), this had not formally been investigated. In order to formally assess any off-device gains in language and cognition the ESPM questionnaire discussed in chapter eight was administered to two of each users immediate caregivers prior to intervention, and then again after six months of regular Easy Speaker use.

The goal of ESPM is to measure improvements in users adaptive behaviour in their everyday environments, as opposed to the purely on-device performance evaluation carried out in chapter nine, or analysis of ARTS tracking data. Specifically, to gauge individuals off-device improvement in communicative and cognitive ability, as compared with their baseline taken prior to Easy

Speaker intervention. Even small improvements in users adaptive behaviour are welcome as they would undoubtedly improve their standard of living and quality of interaction with others. It must be remembered that although behavioural inventories use domains such as memory etc... they are not attempting to measure memory in the same essence as a clinical memory test might. Instead they rate behaviours which would require a memory component to be carried out successfully. Such behavioural measures are commonly used as 'pure' tests of memory and cognition are difficult, if not impossible, to administer to a Learning Disabled non-vocal group.

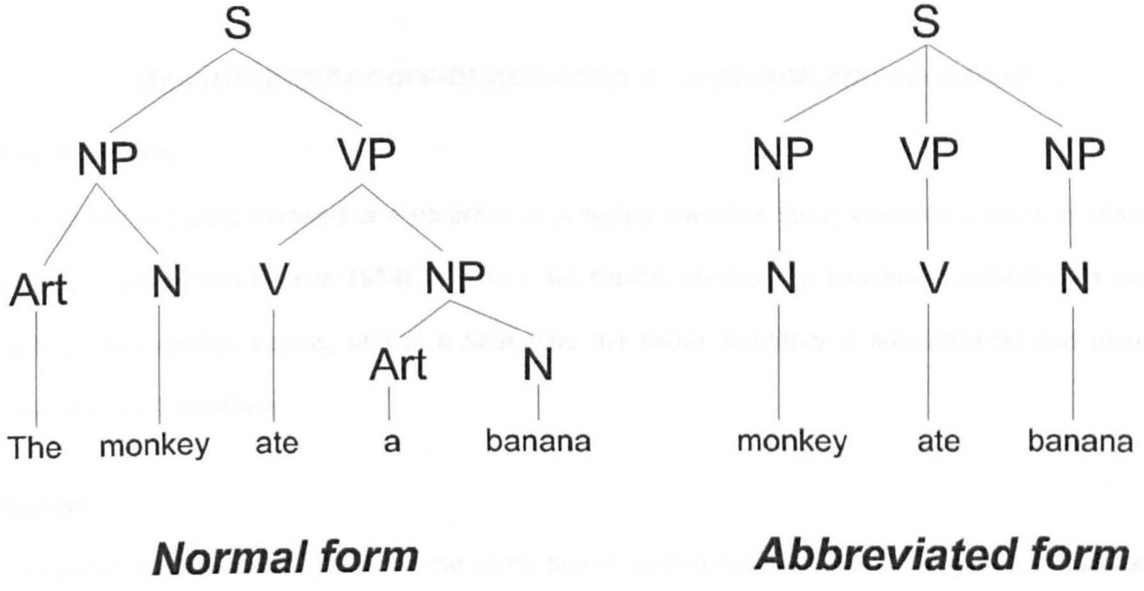
Both pre- and post-intervention sections of the ESPM contained 69 relevant items taken from The Adaptive Behaviour Scales (ABS) and The Functional Performance Record 16 (FPR16), and was subdivided into six domains. All analyses would also be sub-divided into these domains :-

❶ Expressive language
❷ Receptive language
❸ Social language development
❹ Initiative
❺ Attention
❻ Memory

It was decided to focus on 'high' level adaptive behaviours as they would be most likely to be representative of the individuals communicative effectiveness in their everyday environment. The hope being that Easy Speaker intervention would be likely to have a positive impact on these adaptive areas, and thus globally on off-device communication.

As discussed in chapter nine, the speech of all users in the current sample was on the borderline of intelligibility. Their linguistic ability was also minimal and they could only communicate their needs to immediate caregivers with varying degrees of effort and success. None of the users in the sample were using Easy Speaker as their primary means of communication, however performance results obtained in chapter nine suggested all were capable of doing so if communication was more limited. In addition, monitoring through ARTS suggested that all could readily assemble multi-symbol utterances which followed 'near normal' grammatical rules and conveyed salient information. That is, 'near normal' in the respect that the information content usually, and rightly, overrides strict grammatical correctness due to issues of constructional speed and functionality. For example :-

Fig 10.1, Phrase structure rules for the phrase "The monkey ate a banana" and "monkey ate banana"



Despite the differences between the phrase structure for the two utterances, both contain the same information semantically. With the abbreviated form more likely to be produced by a user utilising a symbol based ACD such as Easy Speaker. With relevance to off-device communication most users used a form of telegraphic speech which invariably broke 'normal' grammatical rules and had a low information content. In some cases phrase structure rules can be broken completely, but still the information content is there, and the listener interprets the utterance correctly. For example, "banana ate monkey", can still be understood as "The monkey ate a banana", as one obviously assumes that banana's can't eat monkey's, therefore it must have been the monkey who ate the banana. The cognitive load incurred during the transformational grammar process is placed upon the listener, rather than being correctly formed at source. This makes strict grammatical analysis of users linguistic output, both on and off-device, relatively meaningless and again suggests that behavioural measures of communicative effectiveness would be more appropriate over the short term.

Ultimately it is hoped that more severely impaired individuals will be able to use Easy Speaker as their primary means of communication and will show similar patterns of skill acquisition and generative ability on-device, and increased communicative effectiveness both on and off-device.

Simply, the skills needed to drive Easy Speaker productively and benefits initiated from secondary gains in language and cognition as this chapter investigates.

MEASUREMENT OF OFF-DEVICE GAINS IN LANGUAGE AND COGNITION

The Paradigm

The paradigm being followed is a variation of a typical baseline study commonly used in clinical settings (Barlow and Hersen 1984). Using a set clinical inventory a baseline is established for a subject, intervention begins, and at a later date the same inventory is administered and results compared with baseline.

Design

A repeated measures design where the same two of each users immediate caregivers would complete the ESPM questionnaire at pre-intervention (week 1 or baseline), and again at post-intervention (week 24). In a similar fashion to chapter 9 a single-subject approach with parallel replication is being applied in lieu of a control group (Barlow and Hersen 1984).

Users

Users consisted of the same two male and two female moderately learning disabled adults who participated in the skill acquisition study discussed in chapter nine. For a full profile of each user please see chapter nine.

Apparatus

An ICL 386 SL notebook computer with external VGA colour monitor and keyboard. A LogiTech Audioman connected to the computers parallel port was used for speech sampling and reproduction. A standard mouse was also used.

A copy of the Easy Speaker Progress Monitoring (ESPM) questionnaire comprising of a pre- and post-intervention sections [see chapter eight for a full discussion].

Procedure

Prior to any Easy Speaker intervention ESPM questionnaires were given to two of each users immediate caregivers, or those who knew the user for at least a year and who maintained regular contact with them, either personally or professionally. Caregivers were instructed to complete the pre-intervention section which dealt with the current level of adaptive functioning in the six domains outlined. The post-intervention section of the questionnaire was physically separate from the pre-intervention section and would only be attached when the former was completed and returned. If any caregiver raised queries regarding either section of the questionnaire, these were dealt with by the author.

Once ESPM questionnaires had been completed users were introduced to the author and to Easy Speaker. The following intervention timetable was then adhered to :-

Fig 10.2, How ESPM questionnaire fitted in with the study timetable

Week No./ session No. (timeframe)	Timeframe description	Weekly sessions	Hours per session	Sessions in timeframe	Total hours in timeframe
1-4	Familiarisation/acclimatisation period	1	2	4	8
5	ESAT tasks week 1	1	2	1	2
6	ESAT tasks week 2	1	2	1	2
7	ESAT tasks week 3	1	2	1	2
8-24	Main body of full study	1	2	17	34

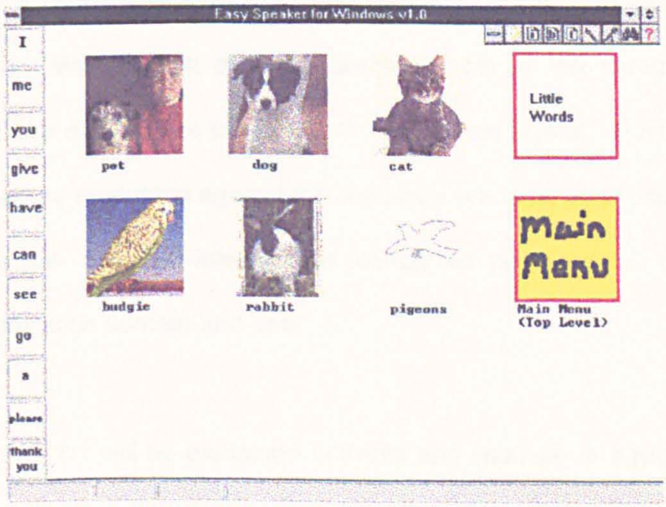
Total number of hours per user over the 24 week study period	48
Total number of hours devoted to all users during each week	8
Total number of hours devoted to users as a whole over the 24 week study period	192

During the four week familiarisation period (weeks 1-4) all users learned to control and make productive use of Easy Speaker. Over the next three weeks (weeks 5-7) users' levels of Easy Speaker skill acquisition and usage was evaluated using the ESAT tasks discussed in chapters eight and nine. Within the main body (weeks 8-24), sessions with Easy Speaker were made up from a mixture of user driven interaction and ad hoc tasks set by the author. Within these sessions the author attempted to bias the clients interactions toward gradually more complex formations. In addition the author freely encouraged the expression of concepts linked to Easy Speaker or to items within the hierarchy. This was done within a metacognitive framework with the intention of enhancing the users own levels of reflectivity and logical reasoning. Where possible the user was prompted to reproduce verbally the utterance they had constructed before being generated as digitised speech.

All users after gaining a rapport with the author and having learned the fundamentals of Easy Speakers operation expressed a desire to have their own pictures scanned and inserted into relevant parts of the hierarchy. Users were also encouraged to record their own "voice" for use with their chosen images and took great pleasure when doing so. All users also expressed a desire to replace music included with the base hierarchy with music more to their own taste. Again this was done with the full involvement of the user and provided another source of pleasure and accomplishment when the results were seen and heard.

Customisation of the original base hierarchy was seen as central to Easy Speaker intervention and was used as an aid to increase the users attention span, and as a reinforcer. When the user began to drift from an interaction with Easy Speaker, the author might suggest that they modified the hierarchy in some way, perhaps by drawing a new image and recording a sound as an accompaniment. One example of customisation is where U1:PS created a screen of pets that he'd had while living at home :-

Fig 10.3, An example of user customisation



With each of these pictures U1 recorded his own voice saying the pets name, e.g. "Sabre the dog". Building utterances which contained references to these newly created pictures was then undertaken. As with any of the utterances constructed during sessions the author pointed out and encouraged the user to correct any construction that was invalid. For example, an utterance that was grammatically flawed was highlighted as were any that were logically incorrect.

During the sessions the goal was to subtlety direct Easy Speaker intervention so that it would be of most benefit to the user. That is, interactions would not be fixed, but instead in flux to enable the client to determine their own therapy destiny. The authors role was to mediate and aid in what the user wanted to achieve. Easy Speaker might then be considered as a communicative and therapeutic environment rather than an isolated tool.

After each user had completed 24 weeks of this type of Easy Speaker intervention the post-intervention section of the ESPM questionnaire was administered to the same two caregivers who had completed the pre-intervention section for that user. Once questionnaires were completed they were analysed accordingly.

OVERVIEW OF THE PRESENTATION OF RESULTS AND DISCUSSION

In order to present the pre- and post-intervention questionnaire data in a coherent way, results are presented for individual users separately. Results for each user will sub-divided into the six domains outlined earlier. Within each of these domains both of the caregivers pre- and post-intervention questionnaire data will be presented in a graphical format. Post-intervention data for each caregiver will then be evaluated against the expected outcome over baseline, i.e. no change. In addition caregivers pre- and post-intervention ratings will be compared to give a measure of interrater agreement for each domain and user.

The results presented for U1 will be explained in depth and shall serve a model for the remaining three users who's results only will be presented. As each users results are presented discussion of those results and suggestions regarding possible implications is made. Once all results have been presented and discussed in turn a more general discussion will consider the findings as a whole.

RESULTS AND DISCUSSION FOR INDIVIDUAL USERS

RESULTS FOR USER 1: PS

As measures of successful outcome are to a large extent dependent on the raters subjective opinion of the clients adaptive behaviour, both pre- and post-intervention, it is relevant to outline their relationship to the user, how long they had known them and their average weekly contact :-

Fig 10.4, Summary of raters relationship to U1

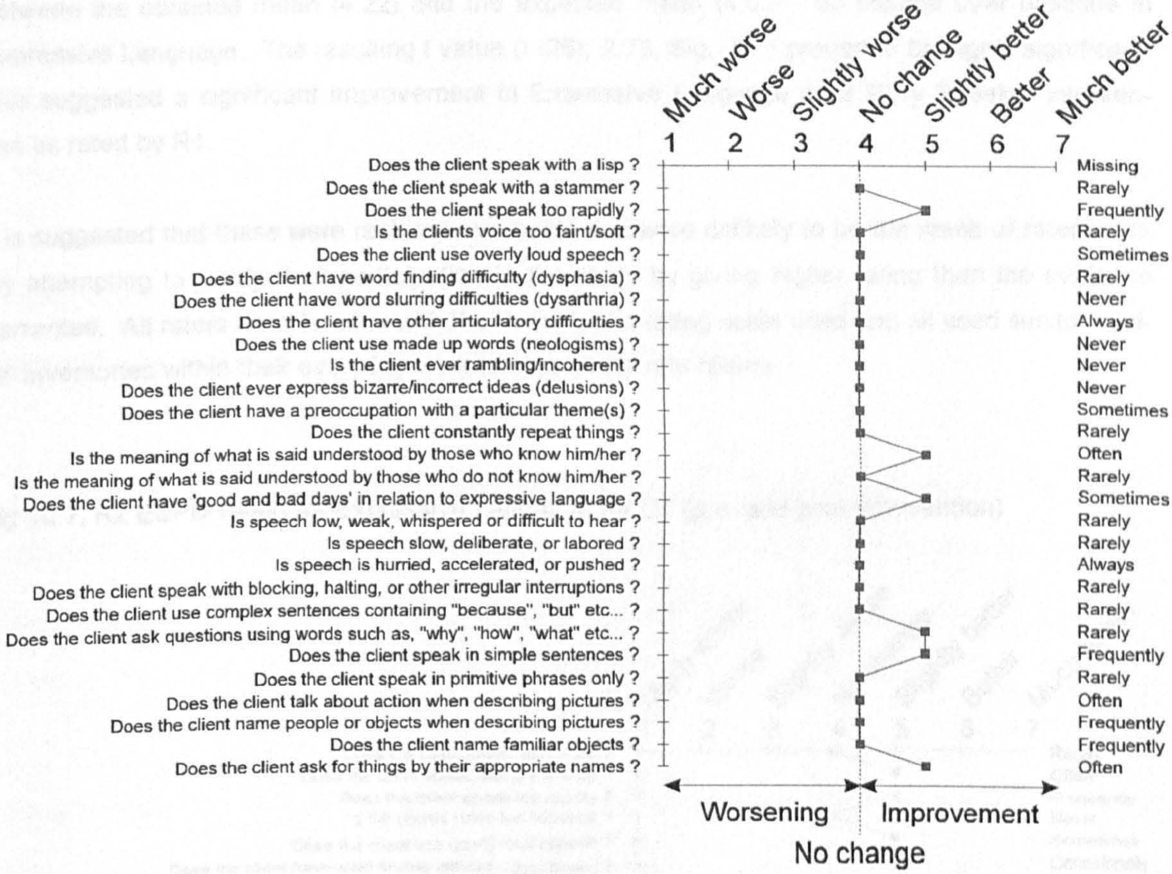
Rater	Relationship to User	How long known User (years)	Ave weekly contact (hours)
R1	Asst. Psychologist	1.5	1
R2	Psychologist	3	0.5

The first domain measured by the ESPM questionnaire at both pre- and post-intervention stages was Expressive Language. Within this domain there were a total of 28 items. In order to simplify and standardise presentation of each raters questionnaire data a graphical method has been adopted.

The graph below shows R1's responses to all 28 items within the expressive language domain at pre- and post-intervention stages. The left axis of the graph lists each questionnaire item in the order that they appeared. The right axis of the graph lists the pre-intervention response to the item made by the rater. Any items on the right with a "Missing" entry represent an item on the questionnaire that the rater has chosen not to complete. The top axis mirrors the semantic differential scale used in the post-intervention questionnaire. Superimposed onto that is a plot of the responses made by the rater at the post-intervention stage. A numerical value has been assigned to each point on the scale, with 4 been taken to represent no change over baseline. The bottom axis simply gives a plain English statement regarding the direction of each point on the graph in relation to the top axis.

EXPRESSIVE LANGUAGE DOMAIN RESULTS FOR U1

Fig 10.5, R1 ESPM rating for Expressive Language for U1 (pre- and post-intervention)



Using item 3 as an example from the graph one might say that "The client 'FREQUENTLY' speaks too rapidly" but after 24 weeks of Easy Speaker intervention they are 'SLIGHTLY BETTER' (capitalised words represent the raters responses to that item). Because the rater has rated them as being "SLIGHTLY BETTER " we can deduce that U1's speech is slightly slower and more controlled that it was prior to Easy Speaker intervention. One can summarise the post-intervention graph by means of the following table :-

Fig 10.6, Summary of post-intervention responses made by R1 in the Expressive Language domain for U1

Mean	SD	Items responded to by rater
4.22	0.42	27

A mean of 4.22 would suggest a slight improvement in Expressive Language after Easy Speaker intervention, as a mean of 4.00 would represent no change, and mean of less than 4.00 a worsening. Although a slight improvement is suggested a single sample t-test was applied to confirm whether this was a significant improvement. This tested whether there was a significant difference between the obtained mean (4.22) and the expected mean (4.00) - no change over baseline in Expressive Language. The resulting t value (t (26), 2.73, Sig. .01) proved to be highly significant. This suggested a significant improvement in Expressive Language after Easy Speaker intervention as rated by R1.

It is suggested that these were real improvements and were unlikely to be the result of raters simply attempting to justify their participation in the study by giving higher rating than the evidence warranted. All raters were familiar with the items in the rating scale used and all used similar clinical inventories within their everyday work to objectively rate clients.

Fig 10.7, R2 ESPM rating for Expressive Language for U1 (pre- and post-intervention)

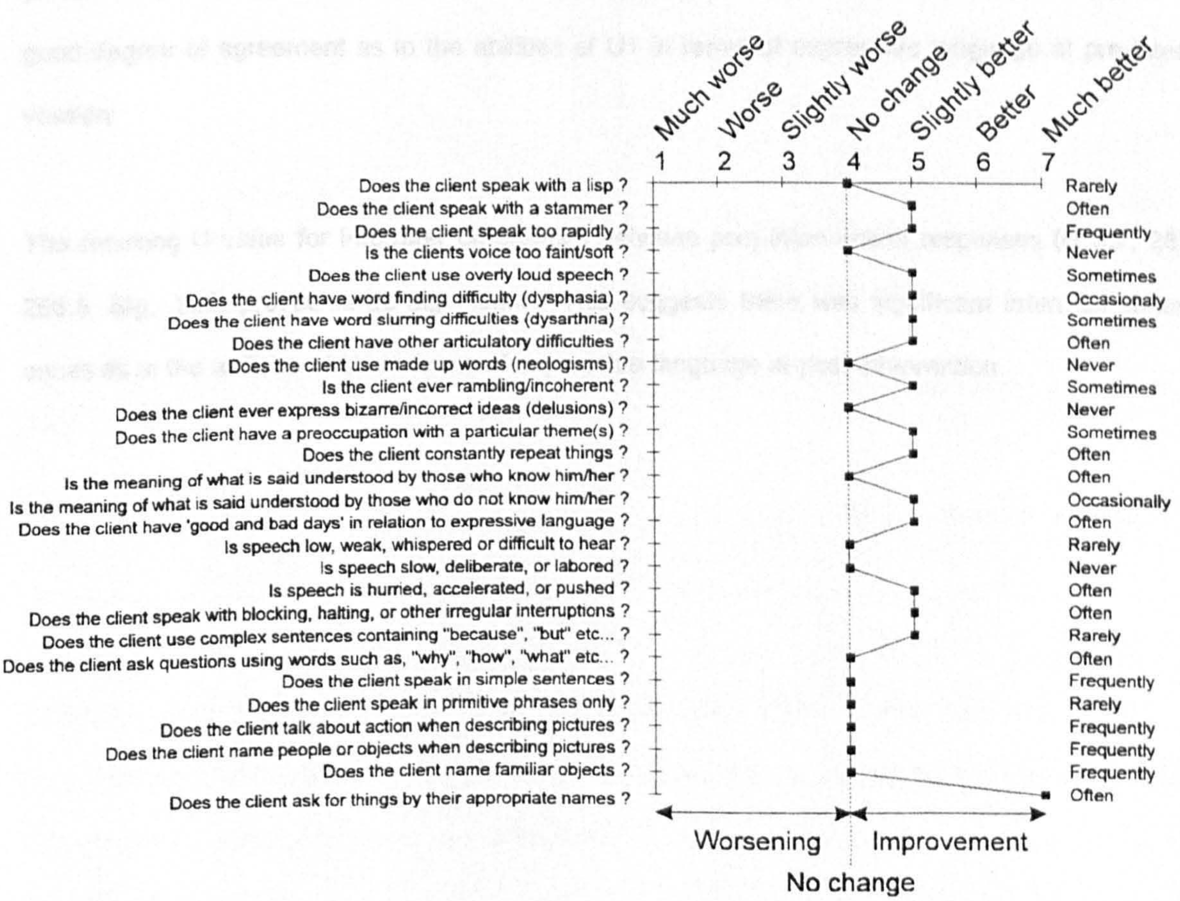


Fig 10.8, Summary of post-intervention responses made by R2 in the Expressive Language domain for U1

Mean	SD	Items responded to by rater
4.61	0.69	28

Again a single sample t-test was applied to test the significance of the improvement suggested by a mean of 4.61. The resulting t value (t (27), 4.69, Sig. .0005) proved to be very highly significant. This suggested a significant improvement in Expressive Language after Easy Speaker intervention as rated by R2.

Interrater differences in pre- and post-intervention responses were assessed by means of an independent t-test and a Mann-Whitney U test respectively. A t-test was applied to pre-intervention data as this tended to be parametric in nature, whereas post-intervention data tended to be less so calling for the application of the non-parametric Mann-Whitney U test.

The resulting t value for interrater differences between pre-intervention responses (t (53), .31, NS) proved to be non-significant. As there were no significant interrater differences this suggests a good degree of agreement as to the abilities of U1 in terms of expressive language at pre-intervention.

The resulting U value for interrater differences between post-intervention responses (U (27, 28), 256.5, Sig. .025) proved to be significant. This suggests there was significant interrater differences as to the abilities of U1 in terms of expressive language at post-intervention.

RECEPTIVE LANGUAGE DOMAIN RESULTS FOR U1

Fig 10.9, R1 ESPM rating for Receptive Language for U1 (pre- and post-intervention)

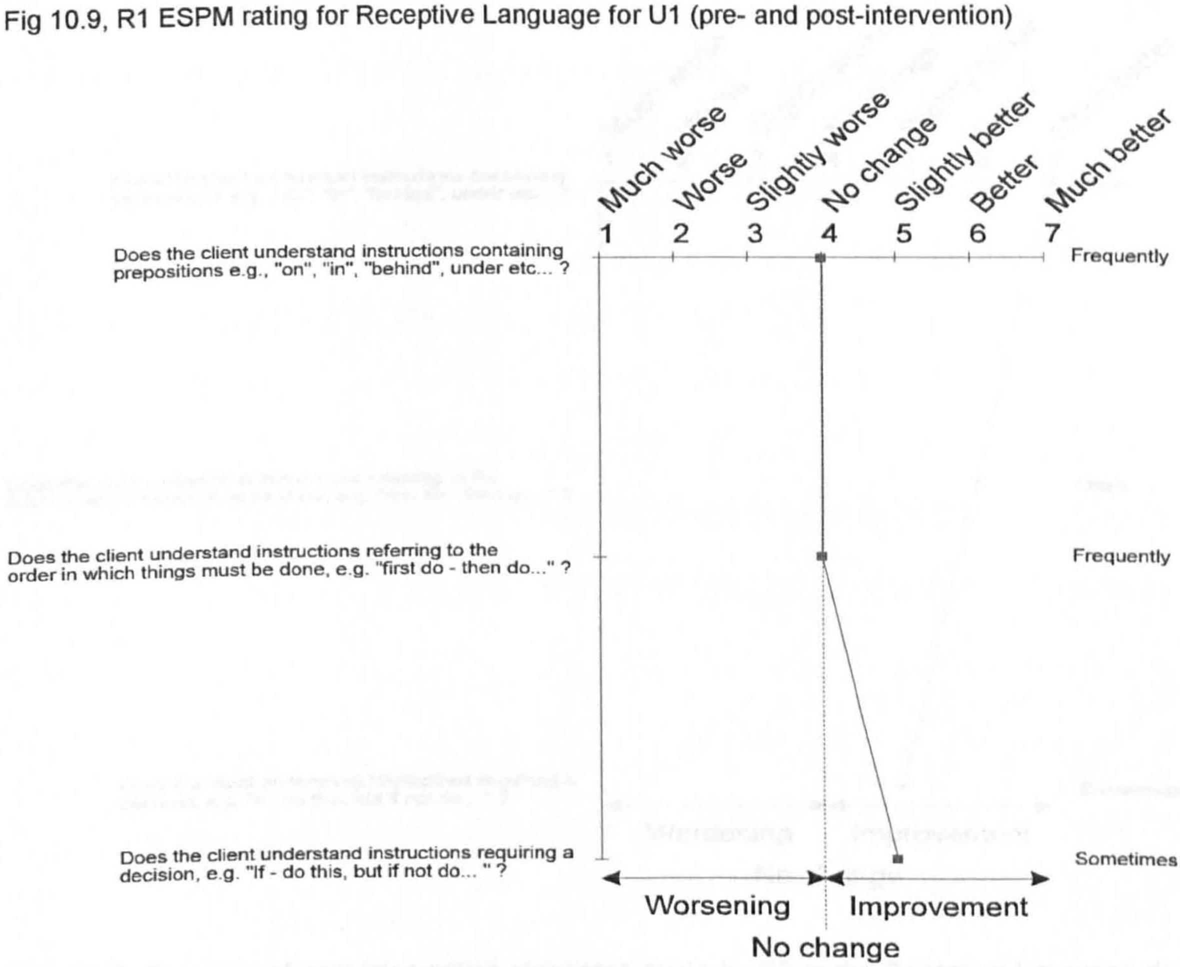


Fig 10.10, Summary of post-intervention responses made by R1 in the Receptive Language domain for U1

Mean	SD	Items responded to by rater
4.33	0.58	3

A single sample t-test was applied to test the significance of the improvement suggested by a mean of 4.33. The resulting t value (t (2), 1.00, NS) proved to be non significant. This suggested there was no significant improvement in Receptive Language after Easy Speaker intervention as rated by R1.

Fig 10.11, R2 ESPM rating for Receptive Language for U1 (pre- and post-intervention)

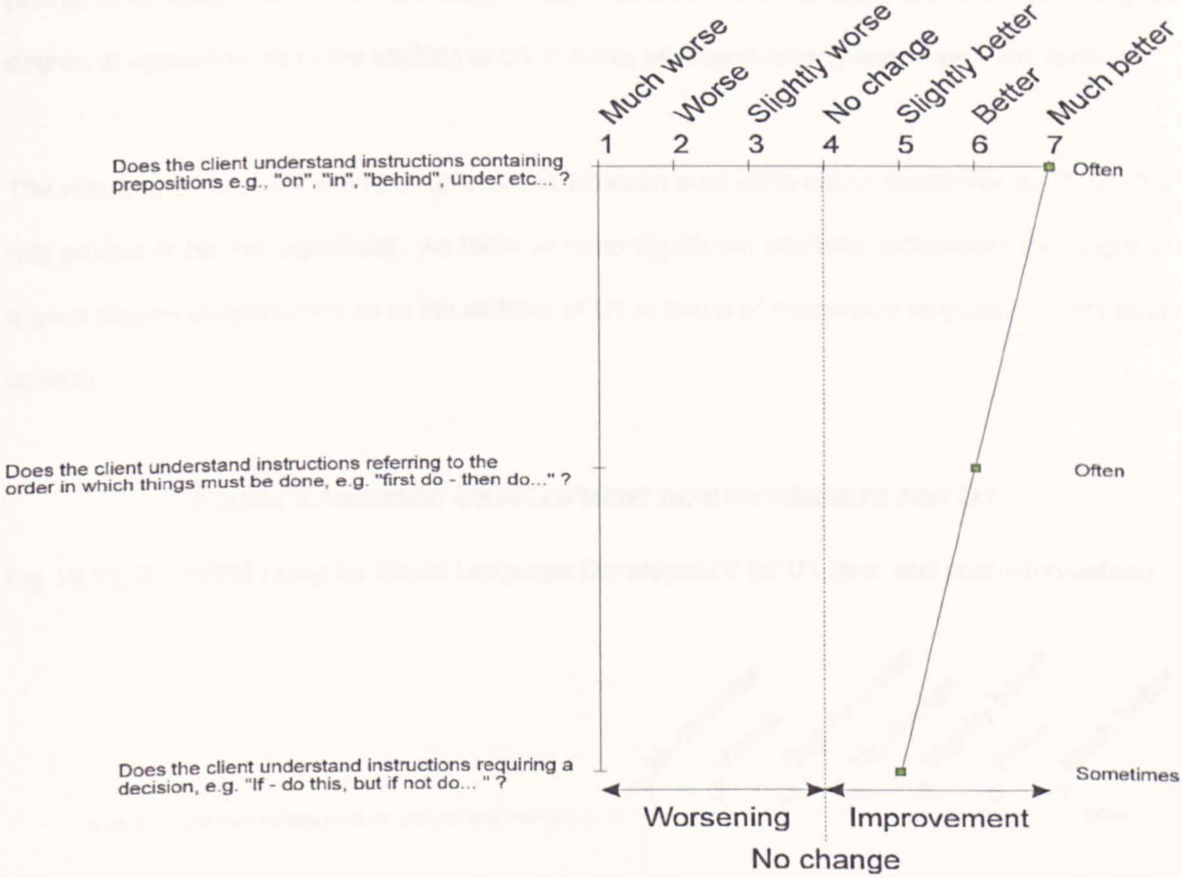


Fig 10.12, Summary of post-intervention responses made by R2 in the Receptive Language domain for U1

Mean	SD	Items responded to by rater
6.00	1	3

A single sample t-test was applied to test the significance of the improvement suggested by a mean of 6.00. The resulting t value (t (2), 3.46, NS) proved to be non significant. This suggested no significant improvement in Expressive Language after Easy Speaker intervention as rated by R2.

Interrater differences in pre- and post-intervention responses were also assessed by means of an independent t-test and a Mann-Whitney U test respectively.

The resulting t value for interrater differences between pre-intervention responses (t (4), .89, NS) proved to be insignificant. As there were no significant interrater differences this suggests a good degree of agreement as to the abilities of U1 in terms of receptive language at pre-intervention.

The resulting U value for interrater differences between post-intervention responses (U (3, 3), 0.5, NS) proved to be non significant. As there were no significant interrater differences this suggests a good degree of agreement as to the abilities of U1 in terms of expressive language at post-intervention.

SOCIAL LANGUAGE DEVELOPMENT DOMAIN RESULTS FOR U1

Fig 10.13, R1 ESPM rating for Social Language Development for U1 (pre- and post-intervention)

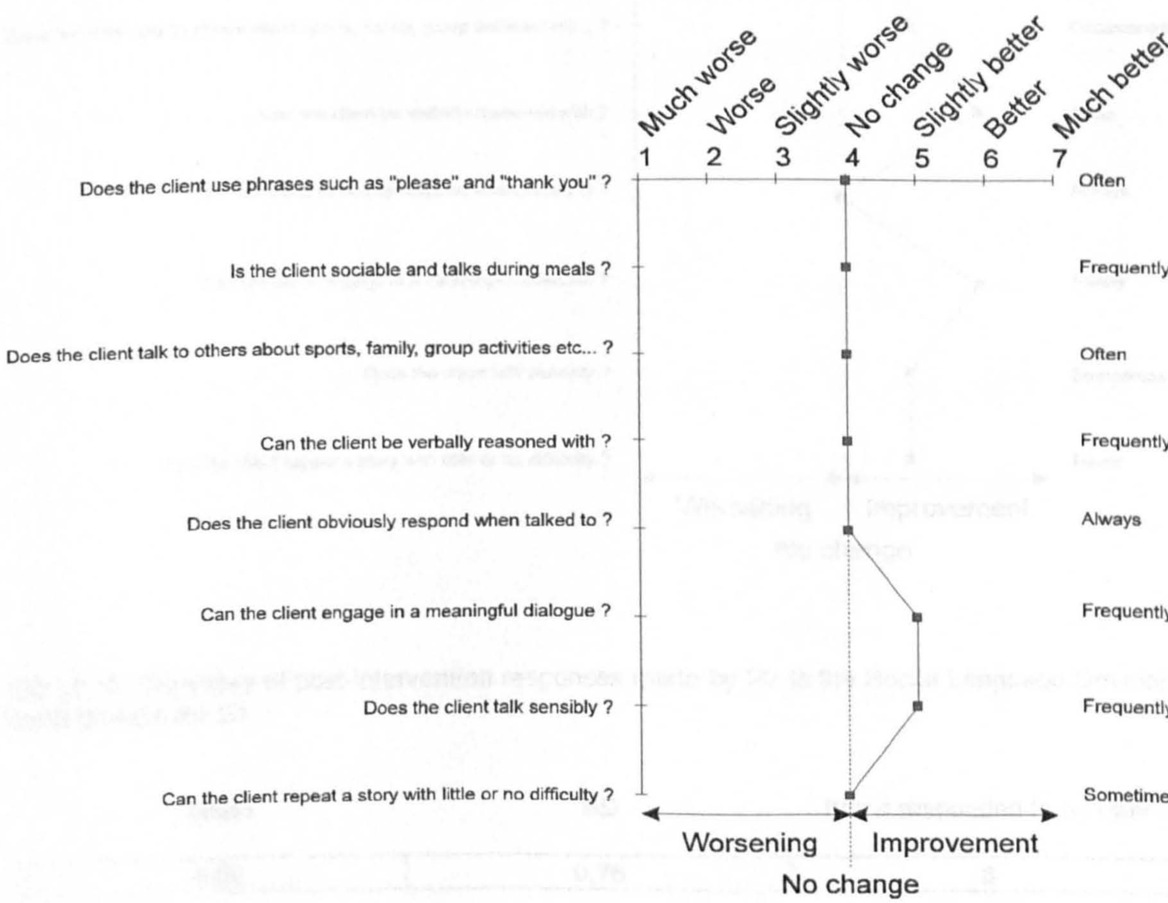


Fig 10.14, Summary of post-intervention responses made by R1 in the Social Language Development domain for U1

Mean	SD	Items responded to by rater
4.25	0.46	8

A single sample t-test was applied to test the significance of the improvement suggested by a mean of 4.25. The resulting t value (t (7), 1.53, NS) proved to be insignificant. This suggested no significant improvement in Social Language Development after Easy Speaker intervention as rated by R1.

Fig 10.15, R2 ESPM rating for Social Language Development for U1 (pre- and post-intervention)

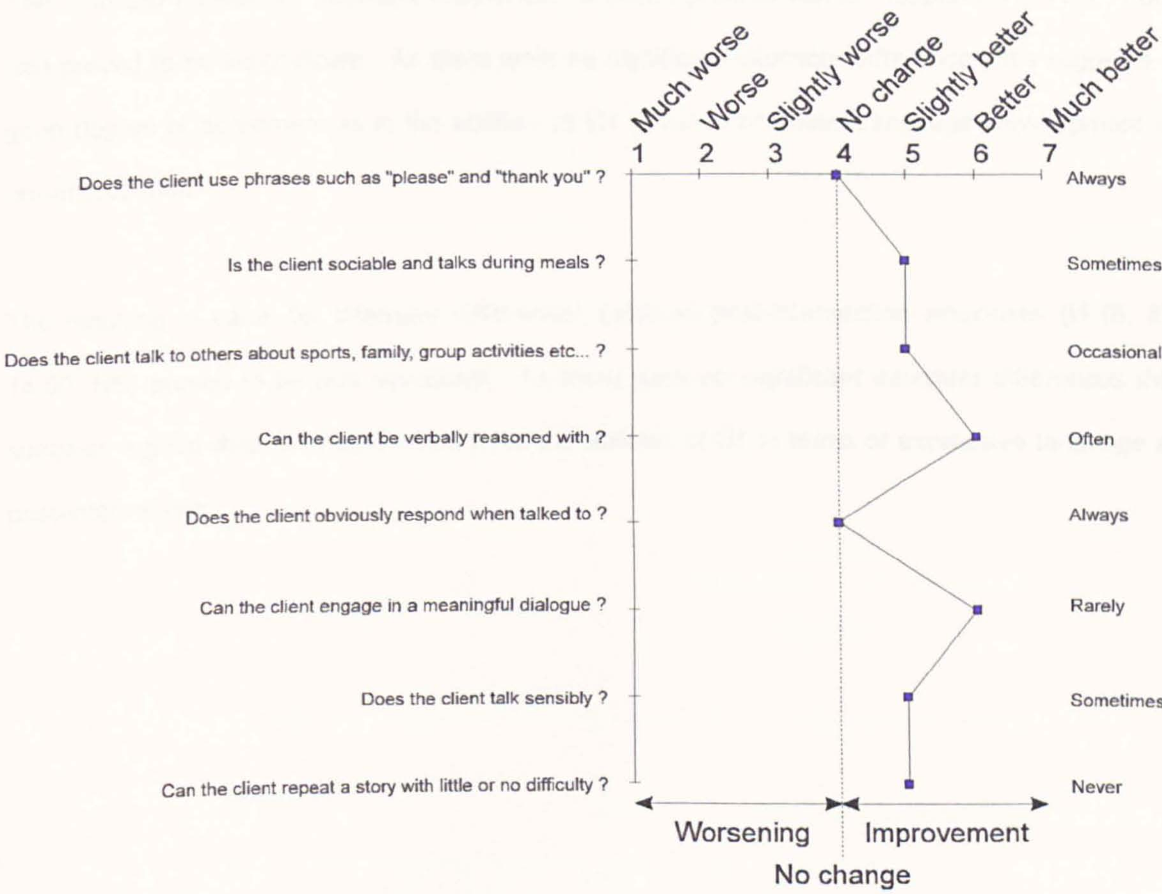


Fig 10.16, Summary of post-intervention responses made by R2 in the Social Language Development domain for U1

Mean	SD	Items responded to by rater
5.00	0.76	8

A single sample t-test was applied to test the significance of the improvement suggested by a mean of 5.00. The resulting t value (t (7), 3.74, .005) proved to be highly significant. This suggested a significant improvement in Social Language Development after Easy Speaker intervention as rated by R2.

Interrater differences in pre- and post-intervention responses were assessed by means of an independent t-test and a Mann-Whitney U test respectively.

The resulting t value for interrater differences between pre-intervention responses (t (14), 1.80, NS) proved to be insignificant. As there were no significant interrater differences this suggests a good degree of agreement as to the abilities of U1 in terms of Social Language Development at pre-intervention.

The resulting U value for interrater differences between post-intervention responses (U (8, 8), 14.00, NS) proved to be non significant. As there were no significant interrater differences this suggests a good degree of agreement as to the abilities of U1 in terms of expressive language at post-intervention.

INITIATIVE DOMAIN RESULTS FOR U1

Fig 10.17, R1 ESPM rating for Initiative for U1 (pre- and post-intervention)

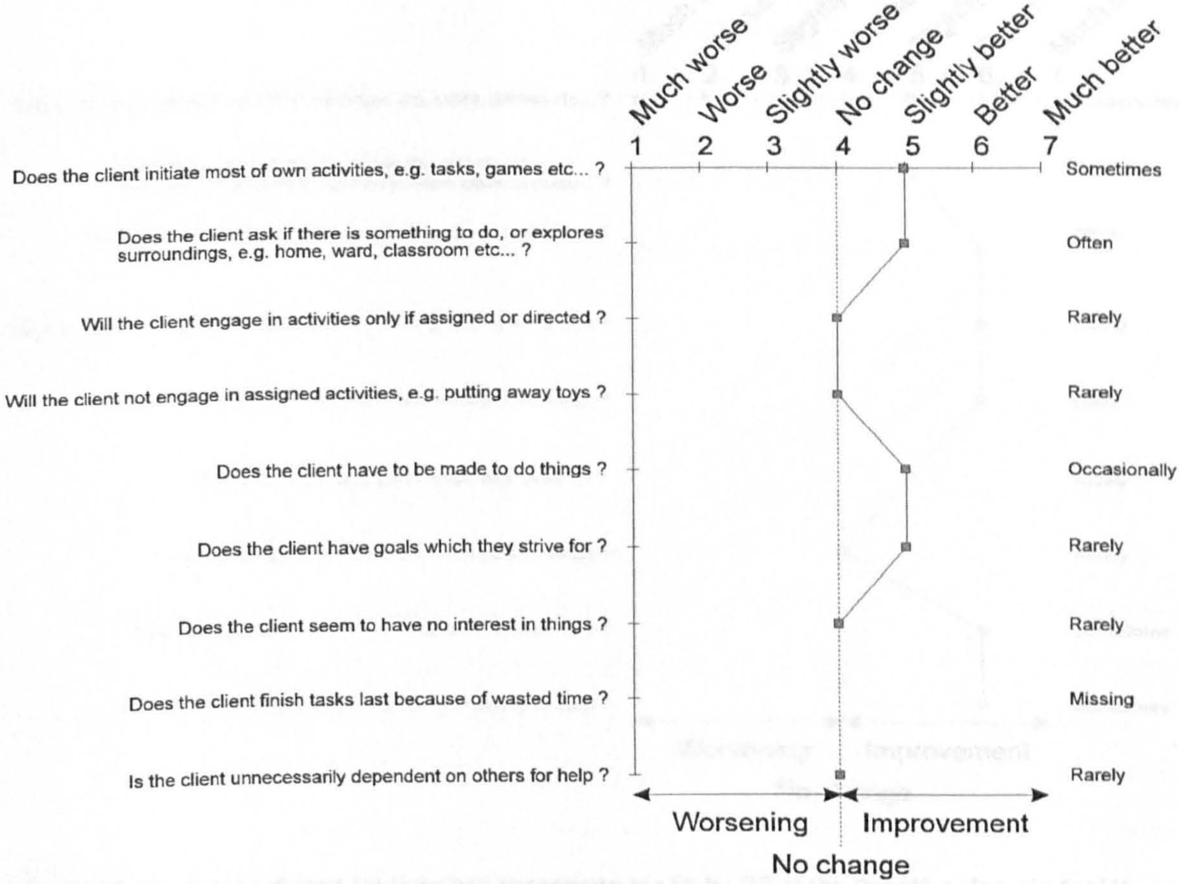


Fig 10.18, Summary of post-intervention responses made by R1 in the Initiative domain for U1

Mean	SD	Items responded to by rater
4.50	0.53	8

A single sample t-test was applied to test the significance of the improvement suggested by a mean of 4.50. The resulting t value (t (7), 2.65, .025) proved to be significant. This suggested a significant improvement in Initiative after Easy Speaker intervention as rated by R1.

Fig 10.19, R2 ESPM rating for Initiative for U1 (pre- and post-intervention)

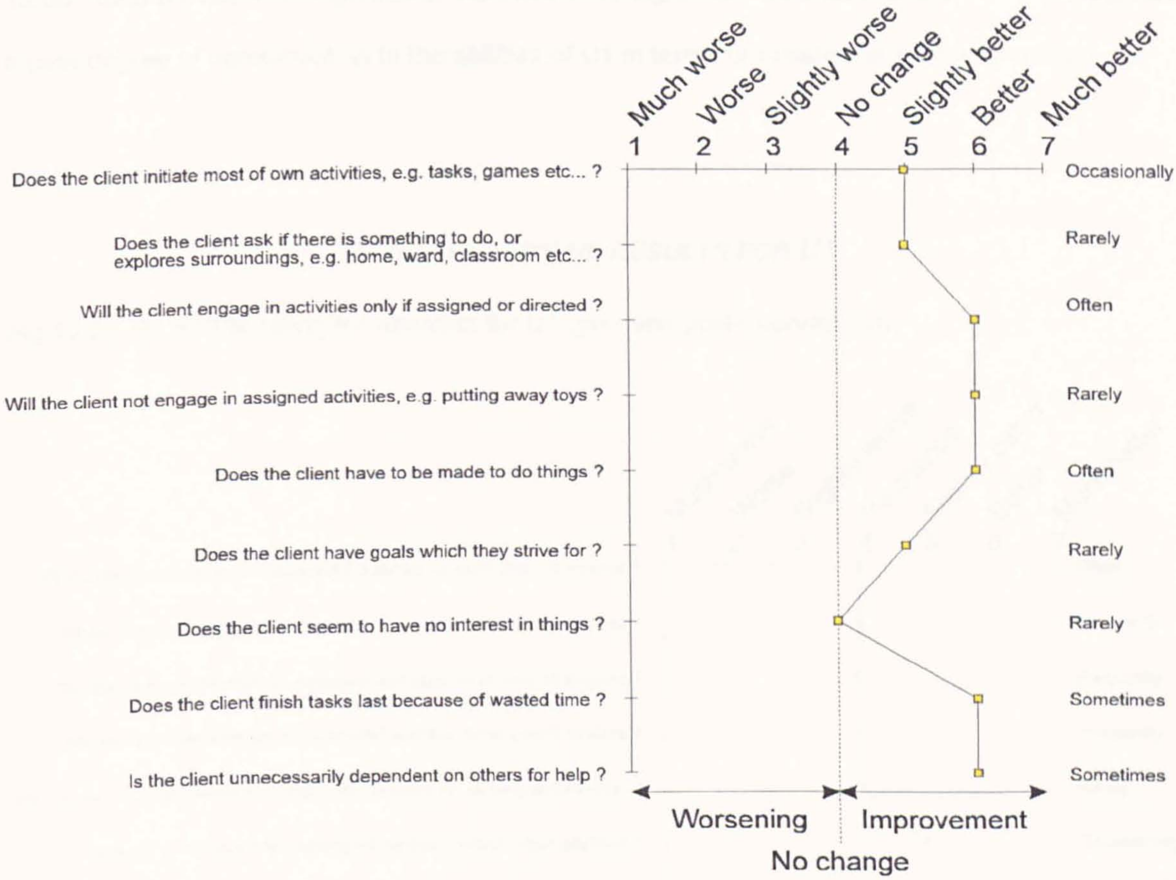


Fig 10.20, Summary of post-intervention responses made by R2 in the Initiative domain for U1

Mean	SD	Items responded to by rater
5.44	0.73	9

A single sample t-test was applied to test the significance of the improvement suggested by a mean of 5.44. The resulting t value (t (8), 5.96, .0005) proved to be very highly significant. This suggested a significant improvement in Initiative after Easy Speaker intervention as rated by R2.

Interrater differences in pre- and post-intervention responses were assessed by means of an independent t-test and a Mann-Whitney U test respectively.

The resulting t value for interrater differences between pre-intervention responses (t (15), 0.78, NS) proved to be insignificant. As there were no significant interrater differences this suggests a good degree of agreement as to the abilities of U1 in terms of Initiative at pre-intervention.

The resulting U value for interrater differences between post-intervention responses (U (8, 9), 12.00, .025) proved to be significant. As there were significant interrater differences this suggests a poor degree of agreement as to the abilities of U1 in terms of Initiative at post-intervention.

ATTENTION DOMAIN RESULTS FOR U1

Fig 10.21, R1 ESPM rating for Attention for U1 (pre- and post-intervention)

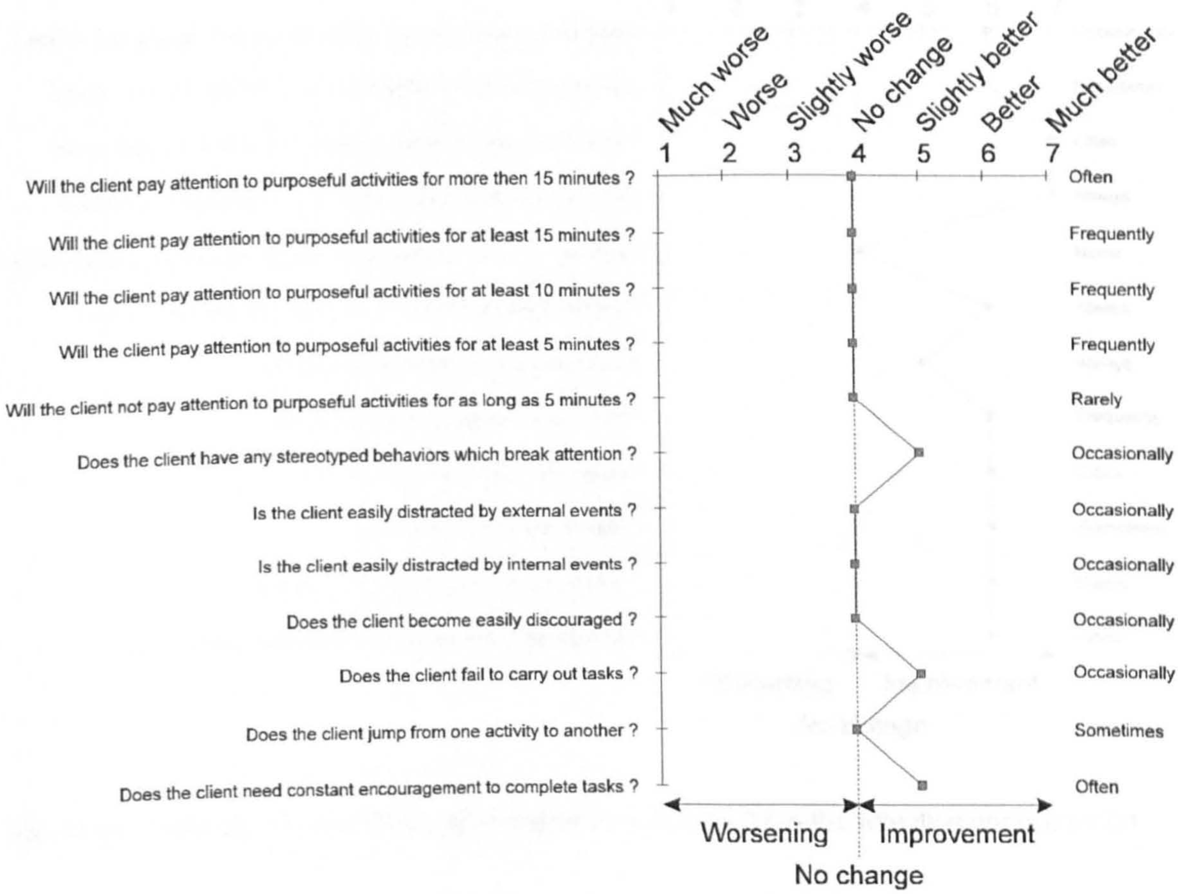


Fig 10.22, Summary of post-intervention responses made by R1 in the Attention domain for U1

Mean	SD	Items responded to by rater
4.25	0.45	12

A single sample t-test was applied to test the significance of the improvement suggested by a mean of 4.25. The resulting t value (t (11), 1.91, .05) proved to be significant. This suggested a significant improvement in Attention after Easy Speaker intervention as rated by R1.

Fig 10.23, R2 ESPM rating for Attention for U1 (pre- and post-intervention)

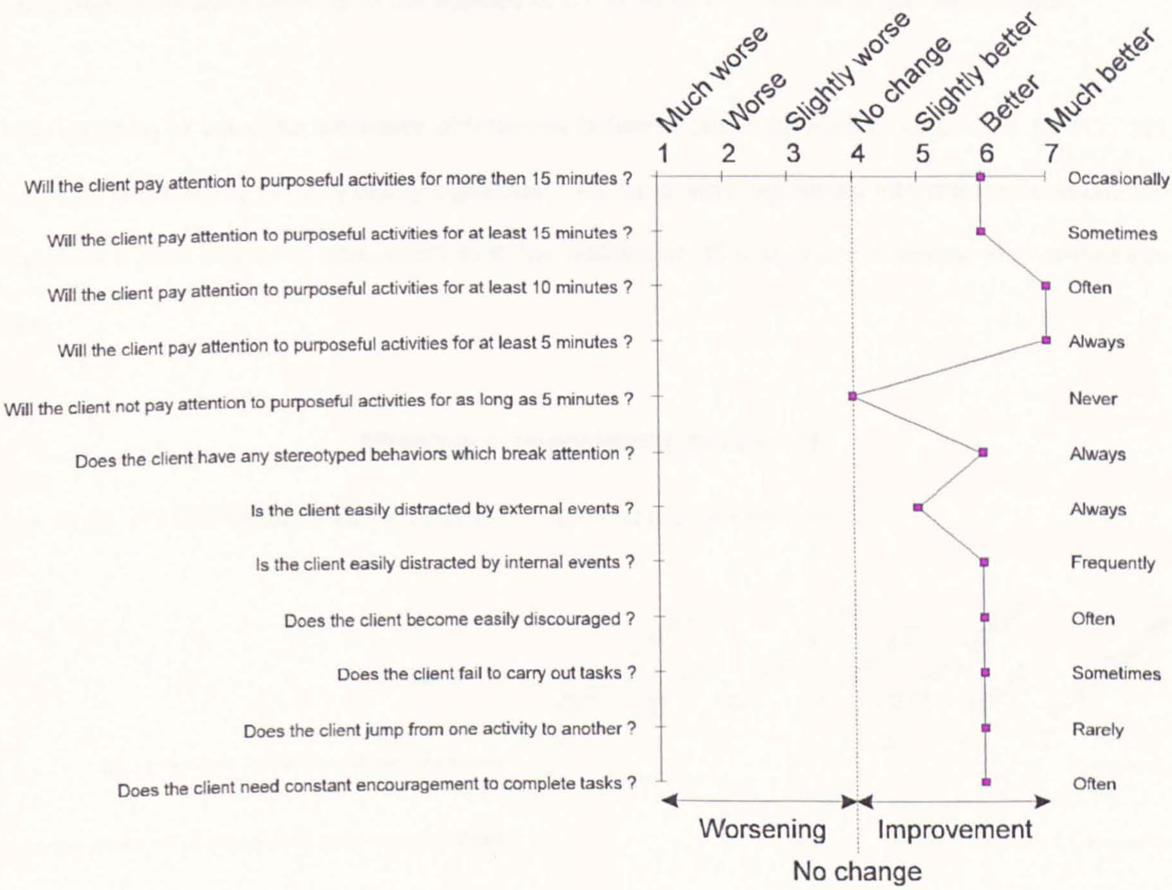


Fig 10.24, Summary of post-intervention responses made by R2 in the Attention domain for U1

Mean	SD	Items responded to by rater
5.91	0.79	12

A single sample t-test was applied to test the significance of the improvement suggested by a mean of 5.91. The resulting t value (t (11), 8.37, .0005) proved to be very highly significant. This suggested a significant improvement in Attention after Easy Speaker intervention as rated by R2.

Interrater differences in pre- and post-intervention responses were assessed by means of an independent t-test and a Mann-Whitney U test respectively.

The resulting t value for interrater differences between pre-intervention responses (t (22), 0.83, NS) proved to be insignificant. As there were no significant interrater differences this suggests a good degree of agreement as to the abilities of U1 in terms of Attention at pre-intervention.

The resulting U value for interrater differences between post-intervention responses (U (12, 12), 9.0, .0001) proved to be very highly significant. As there were significant interrater differences this suggests a poor degree of agreement as to the abilities of U1 in terms of Attention at post-intervention.

MEMORY DOMAIN RESULTS FOR U1

Fig 10.25, R1 ESPM rating for Memory for U1 (pre- and post-intervention)

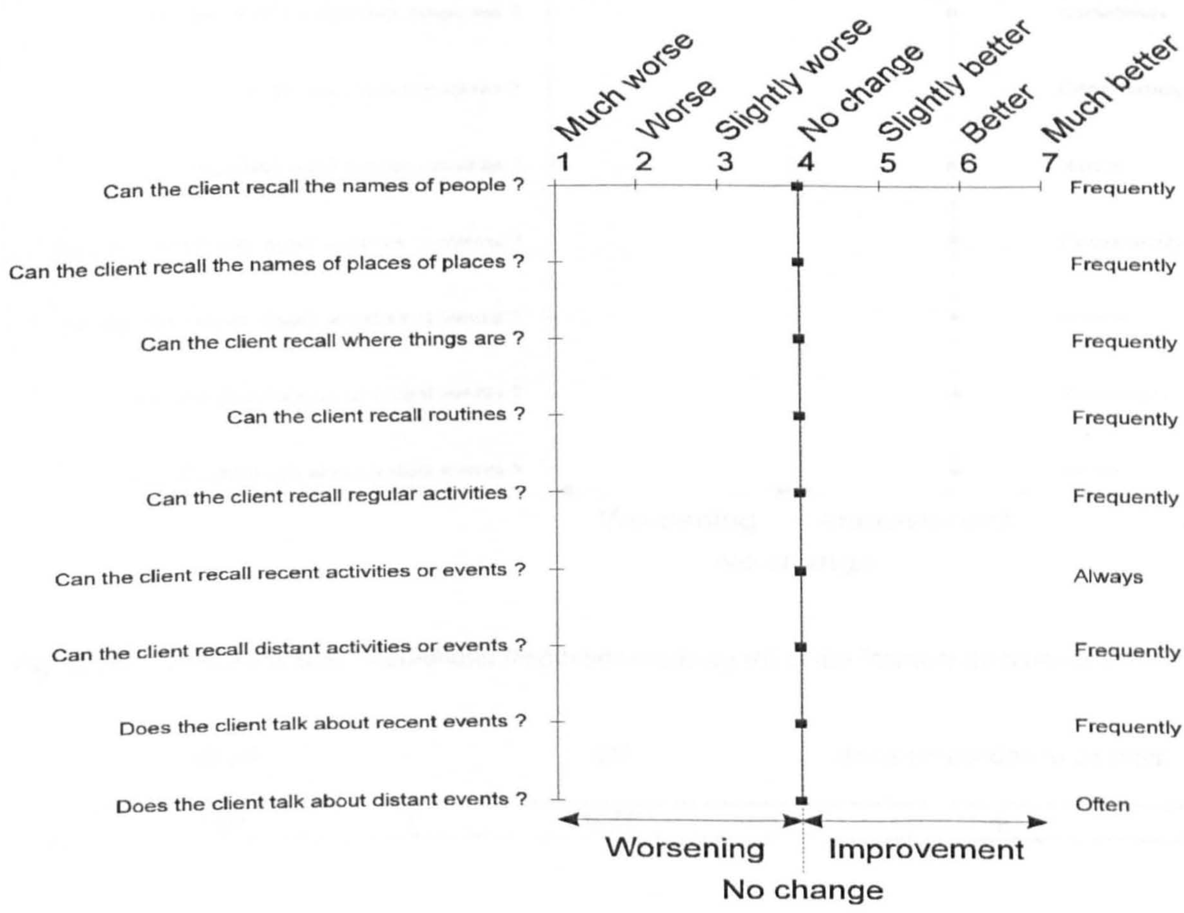


Table 11.26, Summary of post-intervention responses made by R1 in the Memory domain for U1

Mean	SD	Items responded to by rater
4	0	9

A single sample t-test could not be applied in this instance as there was a Standard Deviation of 0. However as a mean of 4 with a SD of 0 was obtained it is clear that there was no significant improvement in Attention after Easy Speaker intervention as rated by R1.

Fig 10.26, R2 ESPM rating for Memory for U1 (pre- and post-intervention)

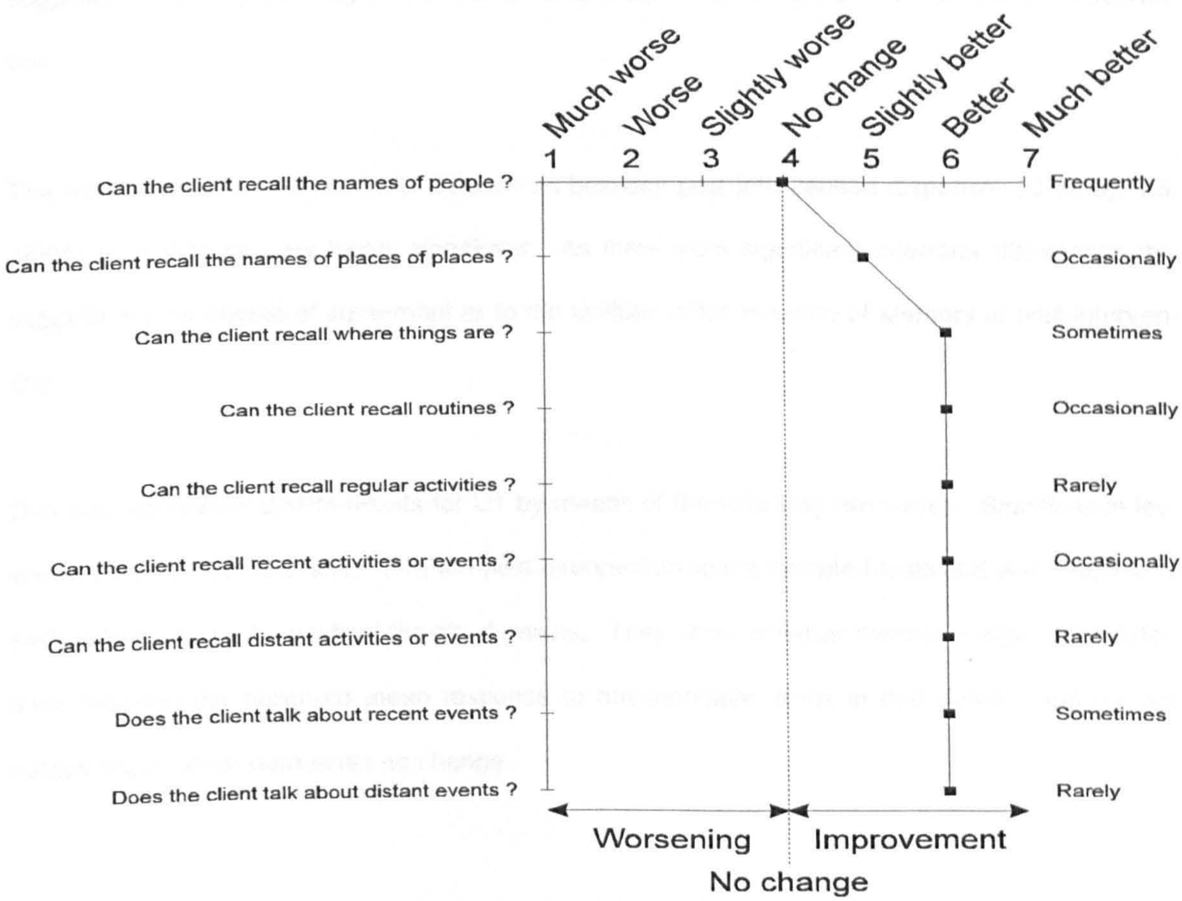


Fig 10.27, Summary of post-intervention responses made by R2 in the Memory domain for U1

Mean	SD	Items responded to by rater
5.67	0.71	9

A single sample t-test was applied to test the significance of the improvement suggested by a mean of 5.67. The resulting t value (t (8), 7.07, .0005) proved to be very highly significant. This suggested a significant improvement in Memory after Easy Speaker intervention as rated by R2.

Interrater differences in pre- and post-intervention responses were assessed by means of an independent t-test and a Mann-Whitney U test respectively.

The resulting t value for interrater differences between pre-intervention responses (t (16), 5.98, .0005) proved to be very highly significant. As there were significant interrater differences this suggests a poor degree of agreement as to the abilities of U1 in terms of Memory at pre-intervention.

The resulting U value for interrater differences between post-intervention responses (U (9, 9), 4.5, .0005) proved to be very highly significant. As there were significant interrater differences this suggests a poor degree of agreement as to the abilities of U1 in terms of Memory at post-intervention.

One can summarise ESPM results for U1 by means of the following two tables. Significance levels on the first table are taken from the post-intervention single sample t-tests that were applied to each raters ratings for each of the six domains. They show whether there is a significant difference between the observed mean response to questionnaire items in that domain and the expected mean which represents no change.

Fig 10.28, A summary of ESPM post-intervention ratings for U1

	Rater R1:PS					Rater R2:SU				
			Sig.	Improvement				Sig.	Improvement	
	t	df	Level	Yes	No	t	df	Level	Yes	No
Expressive Language	2.73	26	0.01	✓		4.69	27	0.0005	✓	
Receptive Language	1.00	2	NS		✓	3.46	2	NS		✓
Social Language Development	1.53	7	NS		✓	3.74	7	0.005	✓	
Initiative	2.65	7	0.025	✓		5.96	8	0.0005	✓	
Attention	1.91	11	0.05	✓		8.37	11	0.0005	✓	
Memory	*	*	*		✓	7.07	8	0.0005	✓	

* Could not be computed due to a Standard Deviation of 0

Significance levels on the first half of the second table are taken from the pre-intervention independent t-tests which were applied to test whether there was a significant difference between the two raters responses in a particular domain. Significance levels of the second half are taken from the post-intervention Mann-Whitney U-tests which were applied to test whether there was a difference between the two raters responses in a particular domain.

Fig 10.29, A summary of ESPM interrater agreement for U1

	Pre Intervention					Post Intervention				
			Sig.	Interrater Agreement				Sig.	Interrater Agreement	
	t	df	Level	Yes	No	U	df	Level	Yes	No
Expressive Language	1.03	53	NS	✓		256.50	27,28	0.025		✓
Receptive Language	0.89	4	NS	✓		0.50	3,3	NS	✓	
Social Language Development	1.80	14	NS	✓		14.00	8,8	NS	✓	
Initiative	0.78	15	NS	✓		12.00	8,9	0.025		✓
Attention	0.83	22	NS	✓		9.00	12,12	0.0001		✓
Memory	5.98	16	0.0005		✓	4.50	9,9	0.0005		✓

DISCUSSION FOR USER 1:PS

Both raters agreed that U1 had improved significantly in three domains, namely; Expressive language, Initiative, and Attention. Interestingly it is within these three areas where U1's disabilities were most marked prior to intervention. Rater R2 also rated improvements in Social language development and Memory, whereas R1 did not. Both agreed that U1 had not improved in terms of Receptive language. U1 was not rated as having worsened on any item in any domain.

Embedded within the Expressive language domain both raters agreed on three sets of items which showed noticeable improvements. U1 was rated as having slowed his rapid speech rate slightly at post-intervention and as a result become more intelligible. Although rated as being only "slightly better", the slowing in U1 speech rate was even commented on by other members of staff who were not involved with the study. U1 was also rated as having good and bad days in relation to Expressive language prior to intervention, whereas at post-intervention he was rated as being "slightly better". This suggests that U1 expressive language became more stable and less erratic in quality. Lastly, both raters rated U1 as showing a greater aptitude for correct naming of things at post-intervention. This may suggest as the author had hoped, that Easy Speaker had strengthened associations between objects and events, and their linguistic label.

Within the initiative domain raters rated U1 as being "better" in terms of initiating most of his own activities. Raters also rated U1 as being more apt to ask if there was something to do, or to explore his surroundings. U1 was also rated as being less likely to have to be made to do things. Raters also rated U1 as having more goals for which he now strove. In sum these improvements suggest that U1 was becoming both more confident in his own abilities and functioning with a higher level of motivation.

In terms of attention raters rated U1's attention as being less likely to be broken by his stereotyped behaviours, such as, finger flicking etc... Indeed these stereotyped behaviours themselves were also reduced; possibly as the result of greater focused attention and increased motivation. Consequently raters rated U1 as also being less likely to fail to carry out tasks. In line with his improvements in initiative U1 was also rated as being less likely to need external encouragement to complete tasks.

Raters could not agree on significant improvements in two domains, namely; Social language development and Memory. Despite lack of interrater agreement rater R2 did rate significant improvements within these areas. Within the Social language development domain R2 rated U1 as being more linguistically sociable, e.g. talking to others during meals, being more likely to engage in meaningful dialogues and to generally talk more sensibly.

Despite U1 already possessing an excellent memory, R2 rated improvements on 9 out of 10 items within the memory domain. However, in U1's case this may be misleading and rather than being pure memory improvements are more likely to be attributed to enhanced confidence and motivation. Simply, that U1 would be more likely to give the correct answer to a question involving memory than remain silent due to lack of confidence in his own memory.

Both raters agreed that there were no significant improvements in Receptive language. This can be attributed to the small number of items within the Receptive language domain, as both raters did actually rate improvements. The implication of such a small number of items will be discussed later.

RESULTS FOR USER 2: AW

Fig 10.30, Summary of raters relationship to U2

Rater	Relationship to user	How long known user (years)	Ave weekly contact (hours)
R1	Asst. Psychologist	1.5	1
R2	Psychologist	3	0.5

EXPRESSIVE LANGUAGE DOMAIN RESULTS FOR U2

Fig 10.31, R1 ESPM rating for Expressive Language for U2 (pre- and post-intervention)

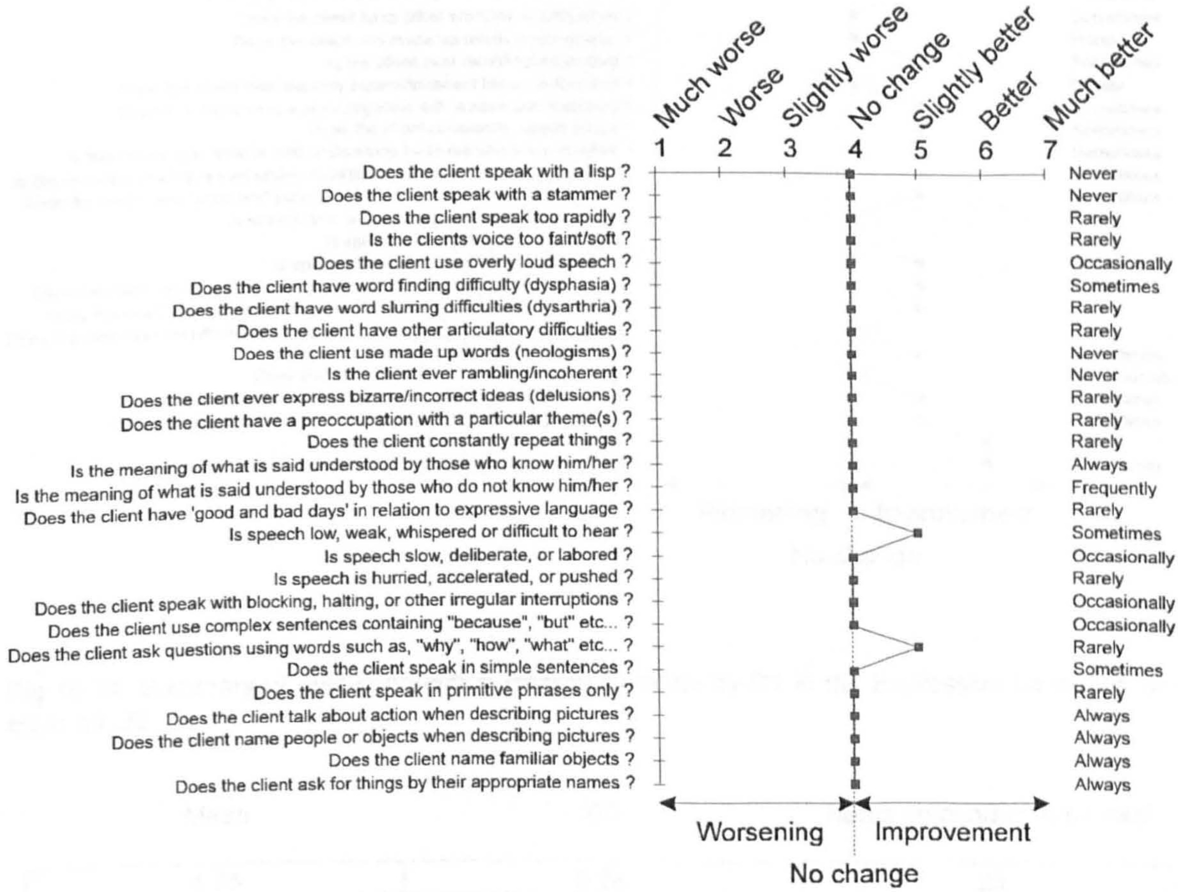


Fig 10.32, Summary of post-intervention responses made by R1 in the Expressive Language domain for U2

Mean	SD	Items responded to by rater
4.07	0.26	28

Fig 10.33, R2 ESPM rating for Expressive Language for U2 (pre- and post-intervention)

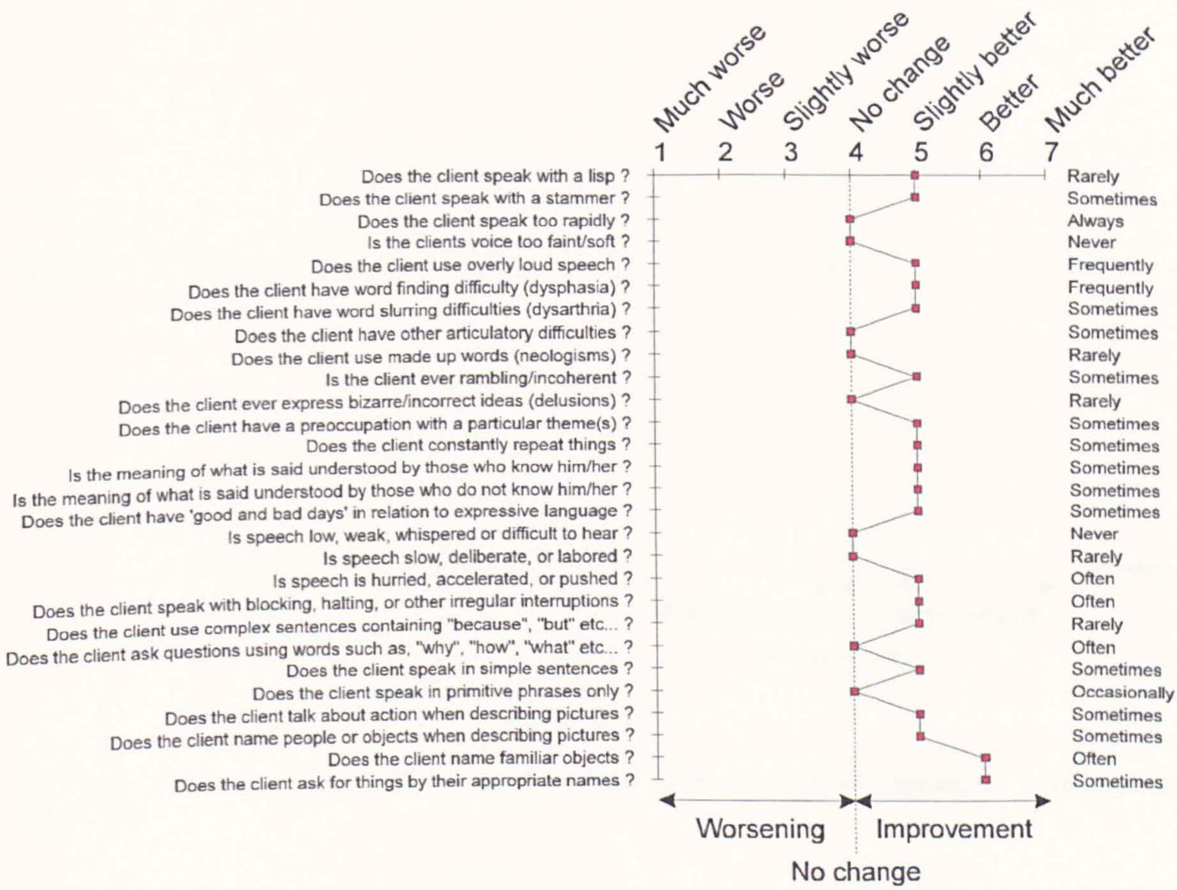


Fig 10.34, Summary of post-intervention responses made by R1 in the Expressive Language domain for U2

Mean	SD	Items responded to by rater
4.75	0.59	28

RECEPTIVE LANGUAGE DOMAIN RESULTS FOR U2

Fig 10.35, R1 ESPM rating for Receptive Language for U2 (pre- and post-intervention)

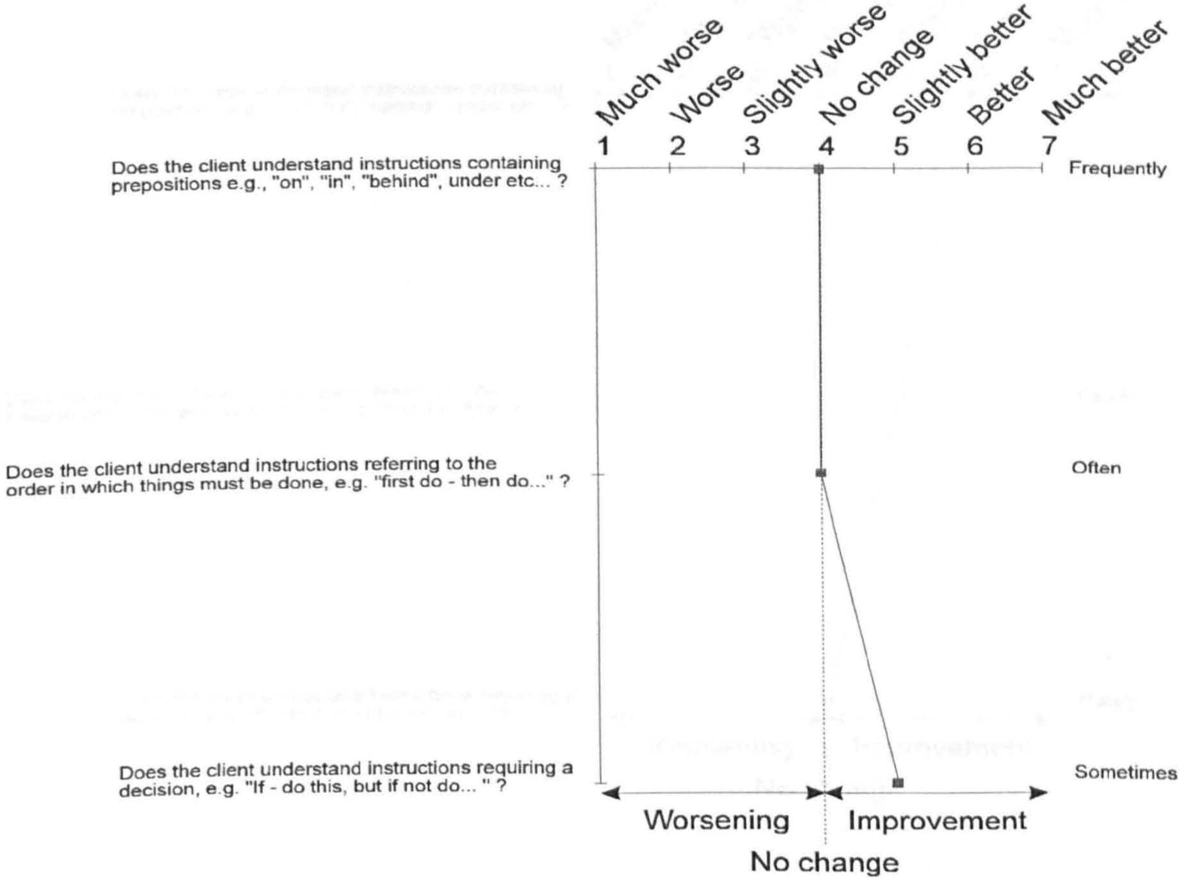


Fig 10.36, Summary of post-intervention responses made by R1 in the Receptive Language domain for U2

Mean	SD	Items responded to by rater
4.33	0.58	3

Fig 10.37, R2 ESPM rating for Receptive Language for U2 (pre- and post-intervention)

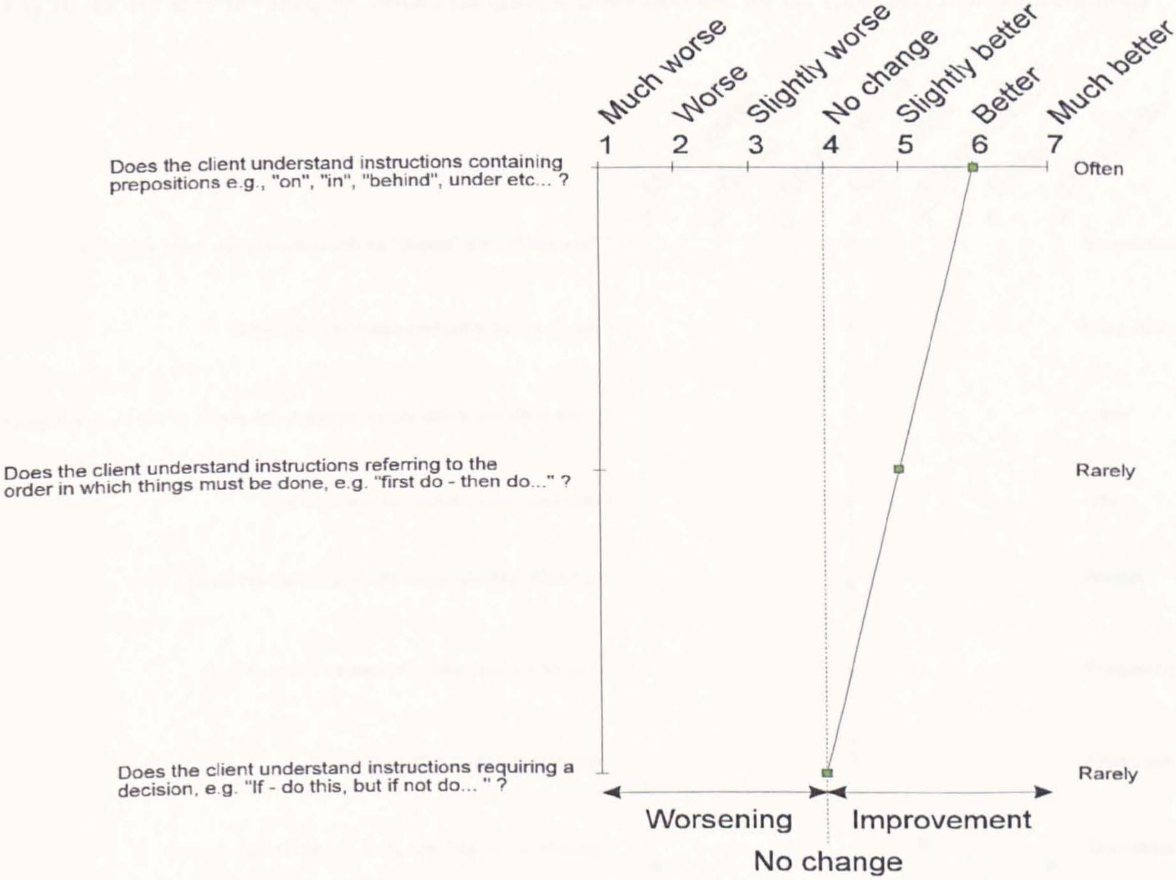


Fig 10.38, Summary of post-intervention responses made by R2 in the Receptive Language domain for U2

Mean	SD	Items responded to by rater
5.00	1	3

SOCIAL LANGUAGE DEVELOPMENT DOMAIN RESULTS FOR U2

Fig 10.39, R1 ESPM rating for Social Language Development for U1 (pre- and post-intervention)

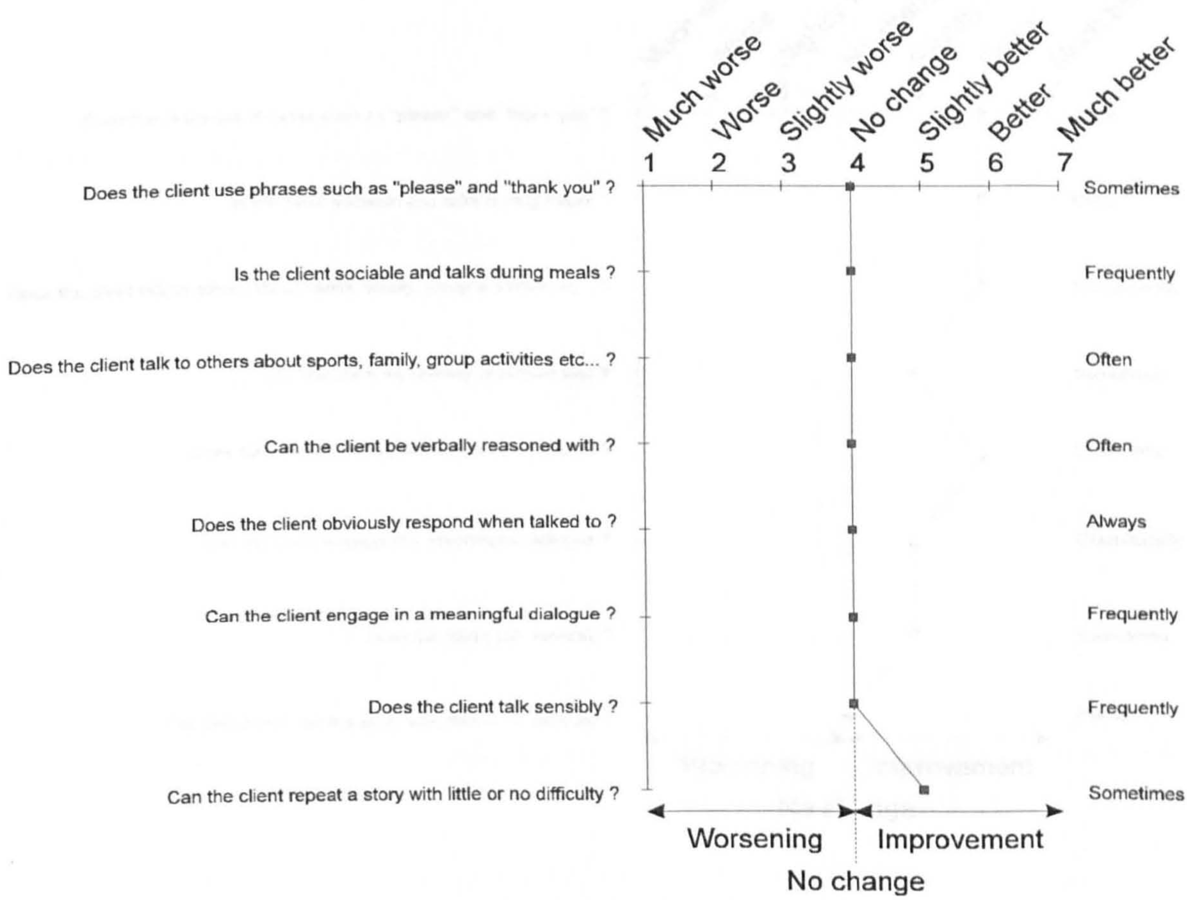


Fig 10.40, Summary of post-intervention responses made by R1 in the Social Language Development domain for U2

Mean	SD	Items responded to by rater
4.13	0.35	8

Fig 10.41, R2 ESPM rating for Social Language Development for U2 (pre- and post-intervention)

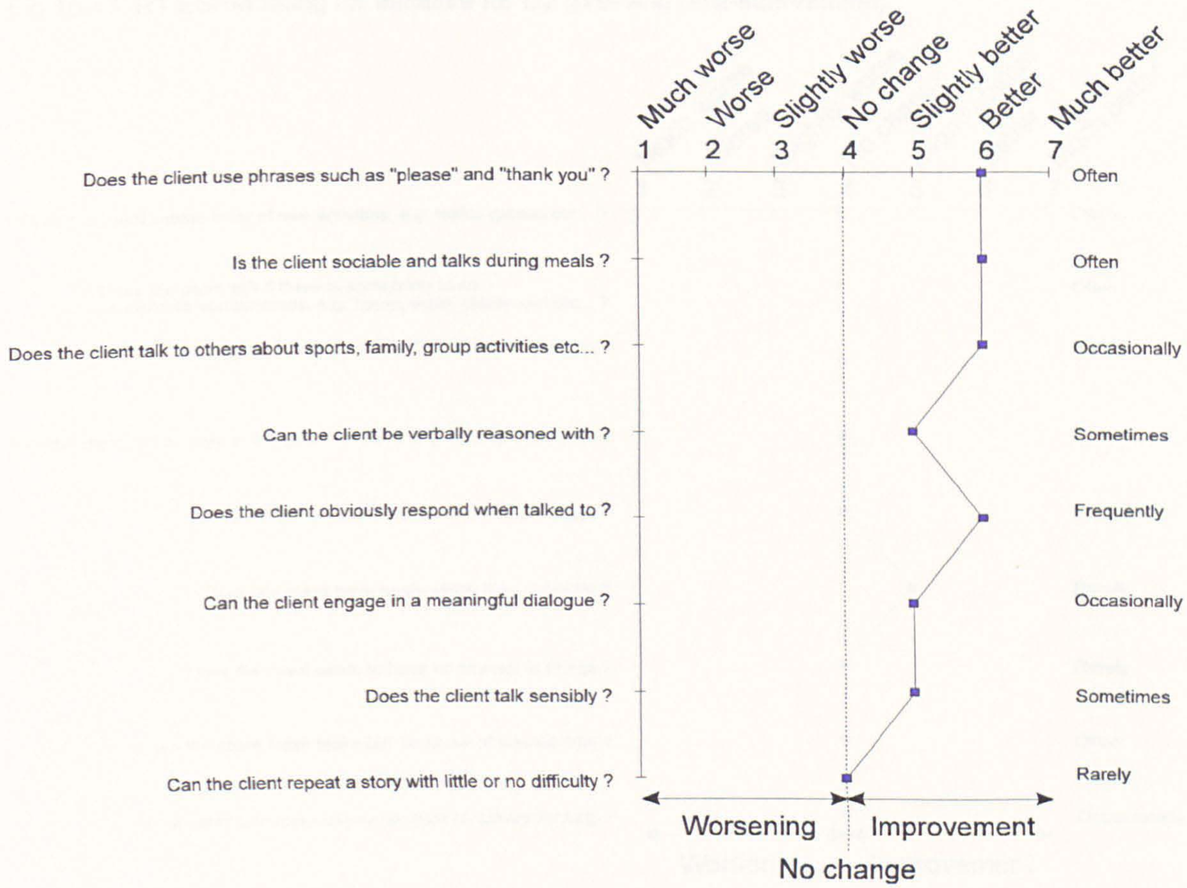


Fig 10.42, Summary of post-intervention responses made by R2 in the Social Language Development domain for U2

Mean	SD	Items responded to by rater
5.38	0.74	8

INITIATIVE DOMAIN RESULTS FOR U2

Fig 10.43, R1 ESPM rating for Initiative for U2 (pre- and post-intervention)

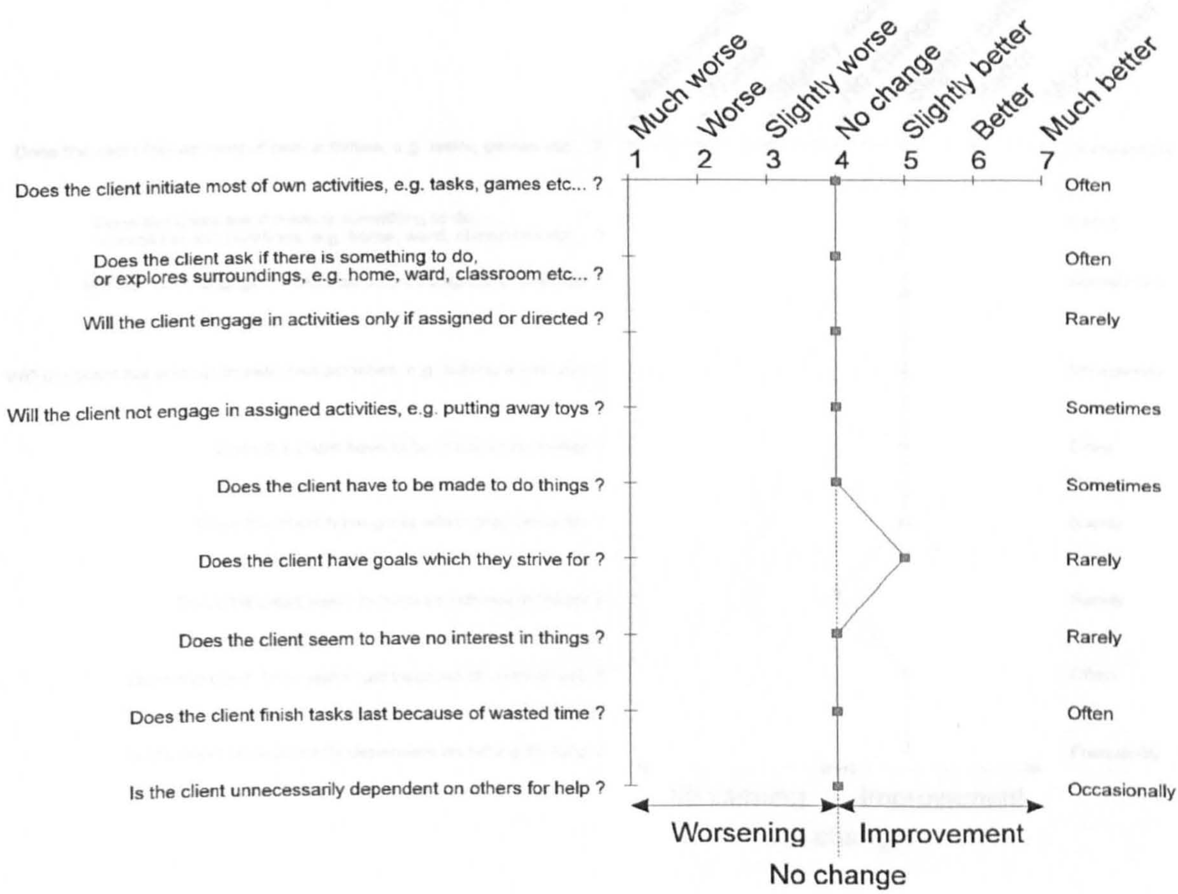


Fig 10.44, Summary of post-intervention responses made by R1 in the Initiative domain for U2

Mean	SD	Items responded to by rater
4.11	0.33	9

Fig 10.45, R2 ESPM rating for Initiative for U2 (pre- and post-intervention)

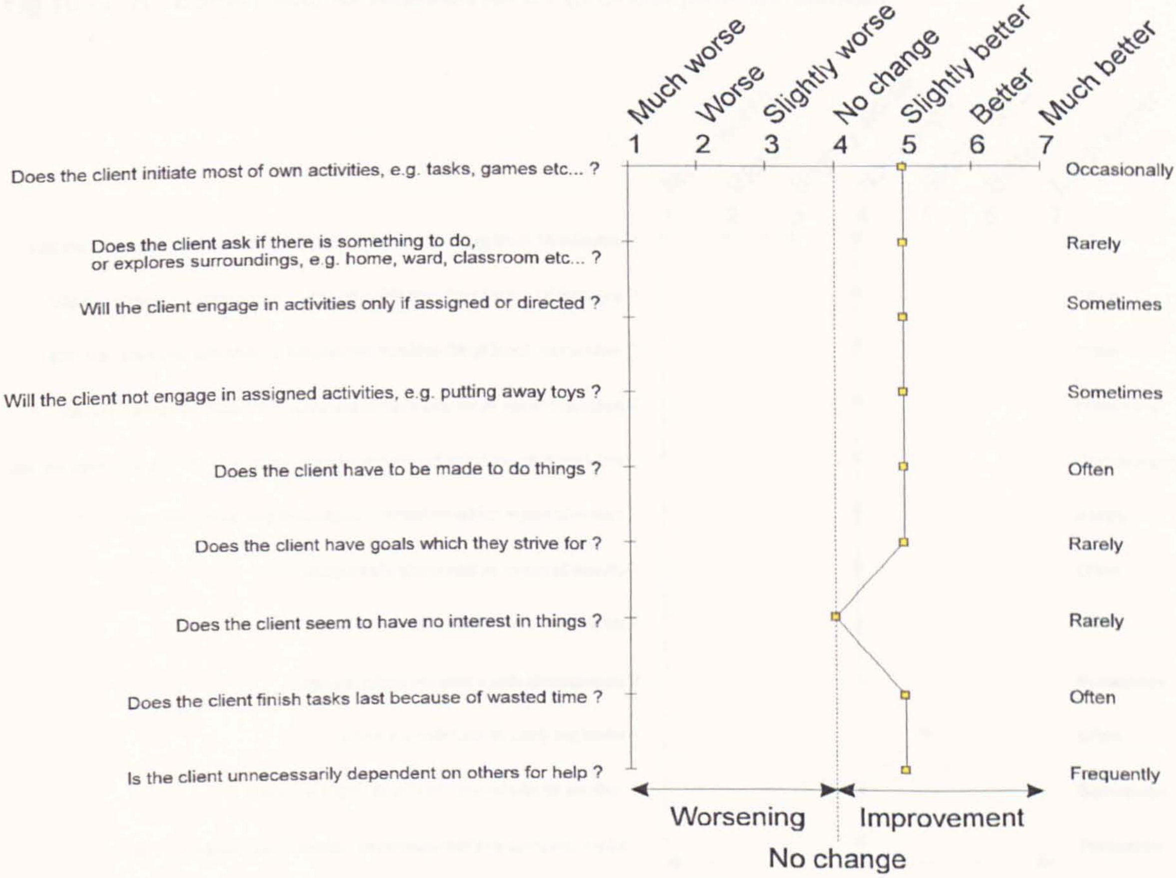


Fig 10.46, Summary of post-intervention responses made by R2 in the Initiative domain for U2

Mean	SD	Items responded to by rater
4.89	0.33	9

ATTENTION DOMAIN RESULTS FOR U2

Fig 10.47, R1 ESPM rating for Attention for U1 (pre- and post-intervention)

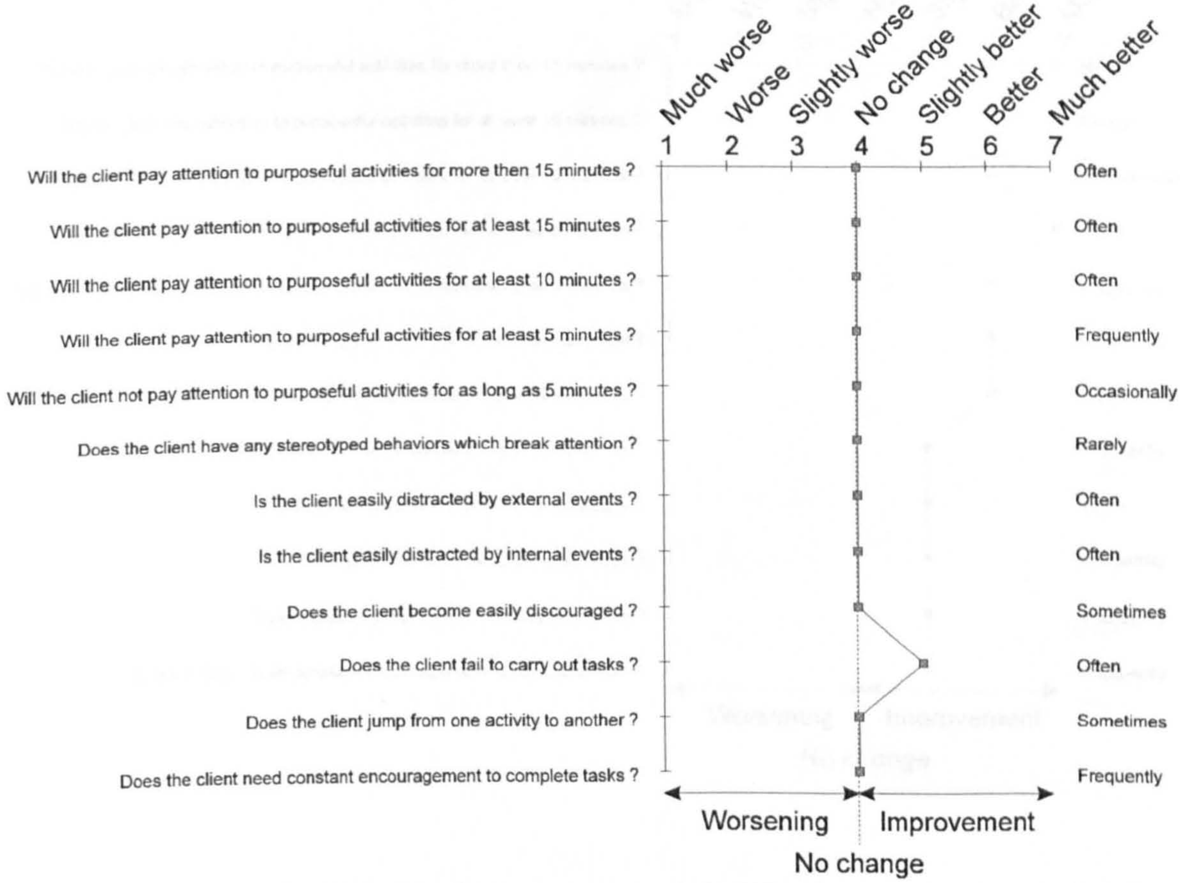


Fig 10.48, Summary of post-intervention responses made by R1 in the Attention domain for U2

Mean	SD	Items responded to by rater
4.08	0.29	12

Fig 10.49, R2 ESPM rating for Attention for U2 (pre- and post-intervention)

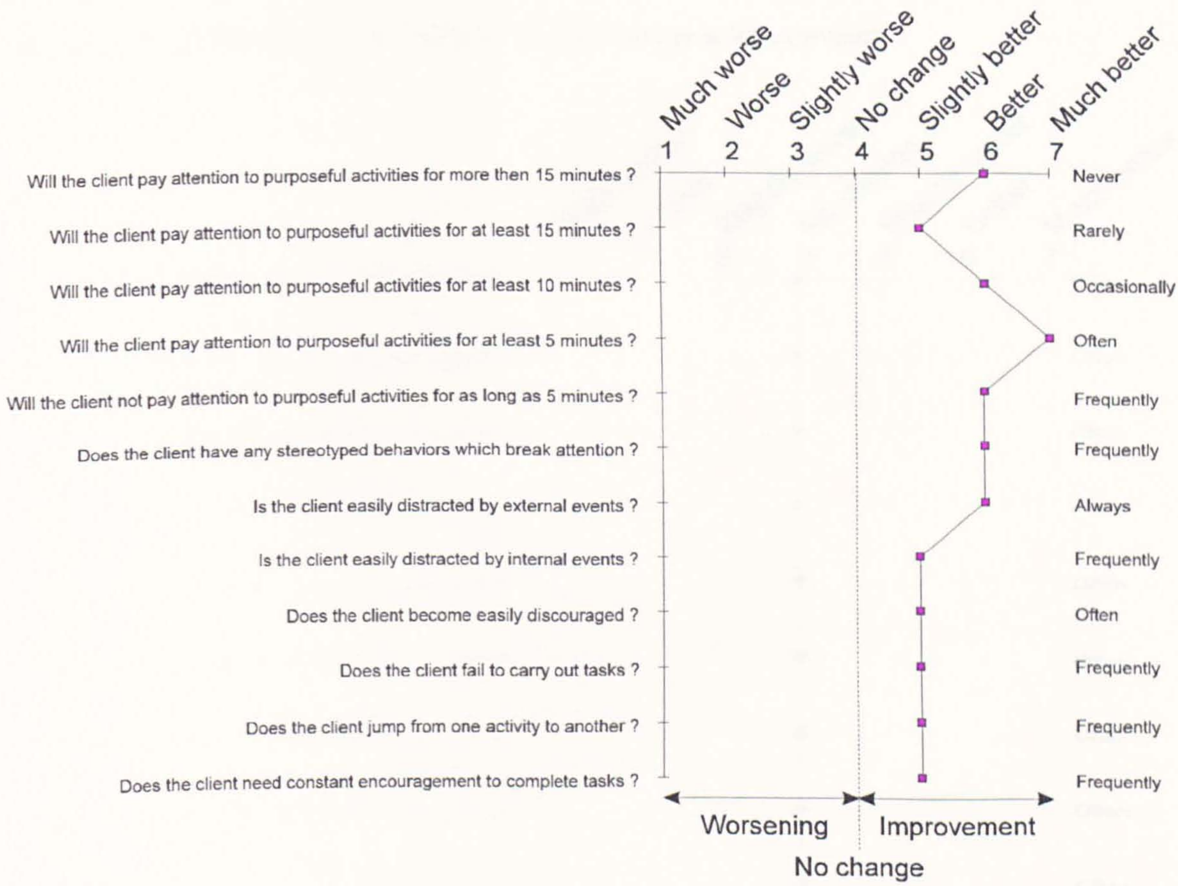


Fig 10.50, Summary of post-intervention responses made by R2 in the Attention domain for U2

Mean	SD	Items responded to by rater
5.58	0.67	12

MEMORY DOMAIN RESULTS FOR U2

Fig 10.51, R1 ESPM rating for Memory for U2 (pre- and post-intervention)

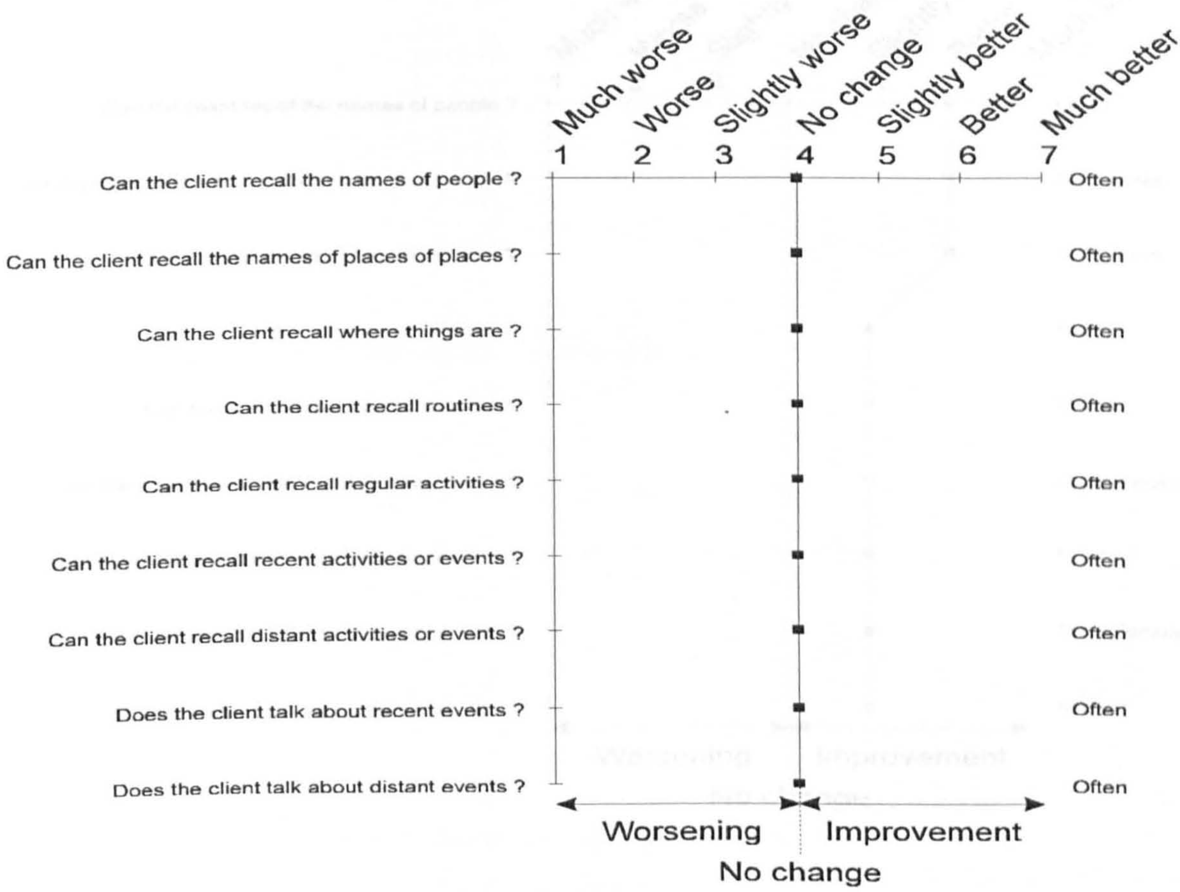


Fig 10.52, Summary of post-intervention responses made by R1 in the Memory domain for U2

Mean	SD	Items responded to by rater
4.00	0	9

Fig 10.53, R2 ESPM rating for Memory for U2 (pre- and post-intervention)

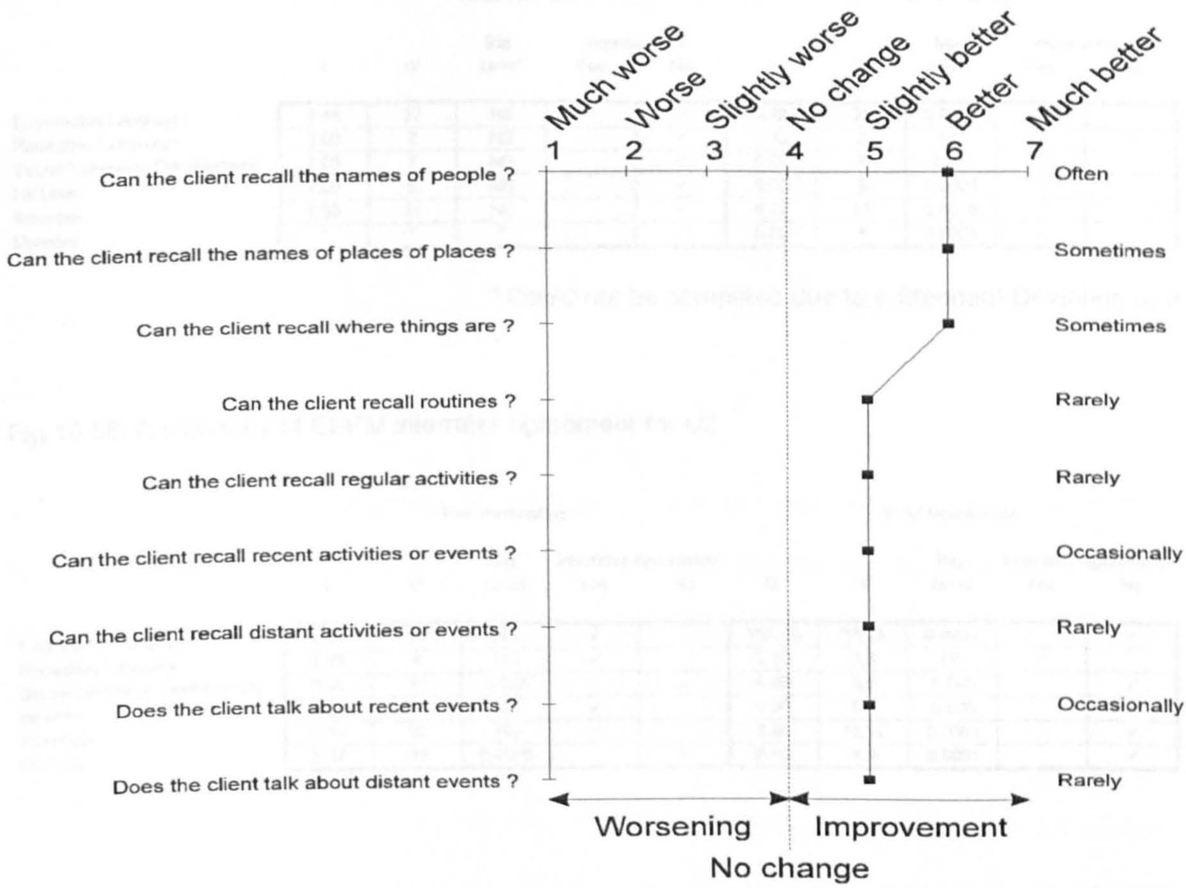


Fig 10.54, Summary of post-intervention responses made by R2 in the Memory domain for U2

Mean	SD	Items responded to by rater
5.33	0.50	9

Fig 10.55, A summary of ESPM post-intervention ratings for U2

	Rater R1:PS					Rater R2:SU				
	t	df	Sig. Level	Improvement Yes	Improvement No	t	df	Sig. Level	Improvement Yes	Improvement No
Expressive Language	1.44	27	NS		✓	6.78	27	0.0005	✓	
Receptive Language	1.00	2	NS		✓	1.73	2	NS		✓
Social Language Development	1.00	7	NS		✓	5.23	7	0.001	✓	
Initiative	1.00	8	NS		✓	8.00	8	0.0005	✓	
Attention	1.00	11	NS		✓	8.20	11	0.0005	✓	
Memory	*	*	*			8.00	8	0.0005	✓	

* Could not be computed due to a Standard Deviation of 0

Fig 10.56, A summary of ESPM interrater agreement for U2

	Pre Intervention					Post Intervention				
	t	df	Sig. Level	Interrater Agreement Yes	Interrater Agreement No	U	df	Sig. Level	Interrater Agreement Yes	Interrater Agreement No
Expressive Language	1.53	54	NS	✓		152.00	28,28	0.0001		✓
Receptive Language	1.73	4	NS	✓		2.50	3,3	NS	✓	
Social Language Development	2.31	14	0.025		✓	6.00	8,8	0.005		✓
Initiative	0.17	16	NS	✓		9.00	9,9	0.005		✓
Attention	0.52	22	NS	✓		3.00	12,12	0.0001		✓
Memory	5.37	16	0.0005		✓	0.00	9,9	0.0001		✓

DISCUSSION FOR USER 2:AW

Both raters could not agree on any domains in which U2 had significantly improved. However, they did agree that he showed no significant improvement within the Receptive language domain.

R1 suggested there had been no overall improvement in any of the six domains, whereas R2 rated a significant improvement in five. Despite R1's lack of agreement and overall ratings of no improvement there were items within each domain in which R1 had rated an improvement. In this light I shall concentrate on those domains in which R2 rated a significant improvement and in tandem highlight those items on which R1 rated an improvement.

Within the Expressive language domain R2 rated improvements in both U2's lisp and pronounced stammer. Improvement was evident in the quietening of his previously overload speech. U2's proneness to dysphasia was also rated as having been reduced. The quality and consistency of U2's speech was also rated as showing improvement. In common with U1 an improved aptitude for correct naming was evident. R1 rated improvements in the degree to which U2 posed "who", "how", "what" ... questions. R1 also commented on the improved quality of U2's speech.

In terms of Social language development R2 rated an improvement in language used in social settings, e.g. being sociable and talking during meals etc... R2 also rated improvements in the extent to which the client would respond to and engage in dialogues, e.g. whether or not U2 could be verbally reasoned with. R1 rated improvement on the degree to which U2 could repeat a story verbally.

U2 was also rated as displaying more initiative, e.g. initiating his own activities, asking for things to do, or exploring his own surroundings. U2 was also rated as being less dependent on others for help and was less apt to be made to have to do things. Both R1 and R2 rated U2 as having more goals for which they now stove.

Improvements on every item of the Attention domain were rated by R2 with, improved attention span, less distractibility, less susceptibility to become discouraged, reduced failure rate when carrying out tasks, better focus on single tasks, and the reduced need for constant encouragement. Both R1 and R2 agreed on a reduced failure rate due to lapses in attention when carrying out tasks.

In common with the Attention domain R2 rated improvements on every item within the Memory domain, whereas R1 rated no improvement on any item. R2 rated him as being more able to recall the names of people and places and to recall where things are in general. He was also rated as being better able to recall regular routines and activities. Recall for both recent and distant events was also rated as having improved, as was his ability to talk about them.

RESULTS FOR USER 3: KC

Fig 10.57, Summary of raters relationship to U3

Rater	Relationship to user	How long known user (years)	Ave weekly contact (hours)
R3*	Care Assistants	1	4 each
R4	Head of Dept	1	0.5

* R3 actually consisted of the combined effort of two care assistants who completed the ESPM questionnaire together for U3

EXPRESSIVE LANGUAGE DOMAIN RESULTS FOR U3

Fig 10.58, R3 ESPM rating for Expressive Language for U3 (pre- and post-intervention)

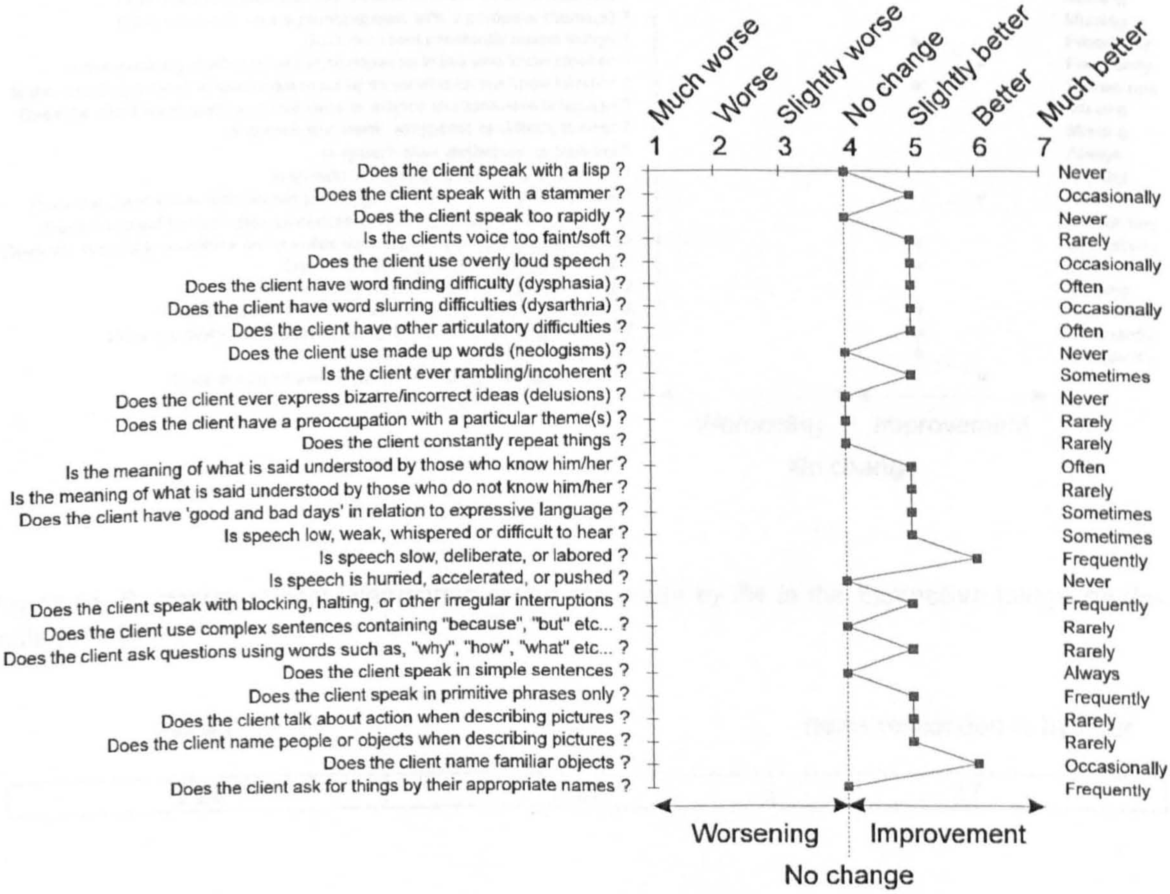


Fig 10.59, Summary of post-intervention responses made by R3 in the Expressive Language domain for U3

Mean	SD	Items responded to by rater
4.71	0.60	28

Fig 10.60, R4 ESPM rating for Expressive Language for U3 (pre- and post-intervention)

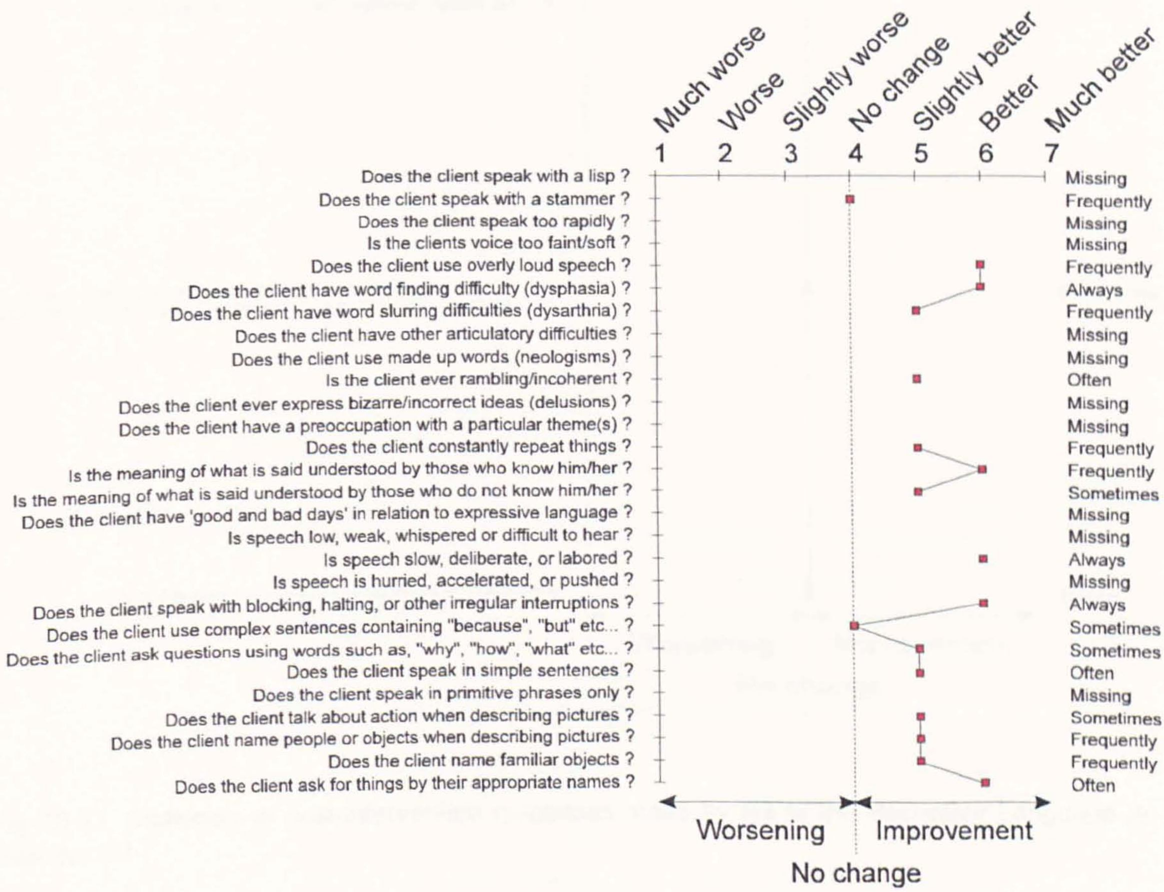


Fig 10.61, Summary of post-intervention responses made by R4 in the Expressive Language domain for U3

Mean	SD	Items responded to by rater
5.24	0.66	17

RECEPTIVE LANGUAGE

Fig 10.62, R3 ESPM rating for Receptive Language for U3 (pre- and post-intervention)

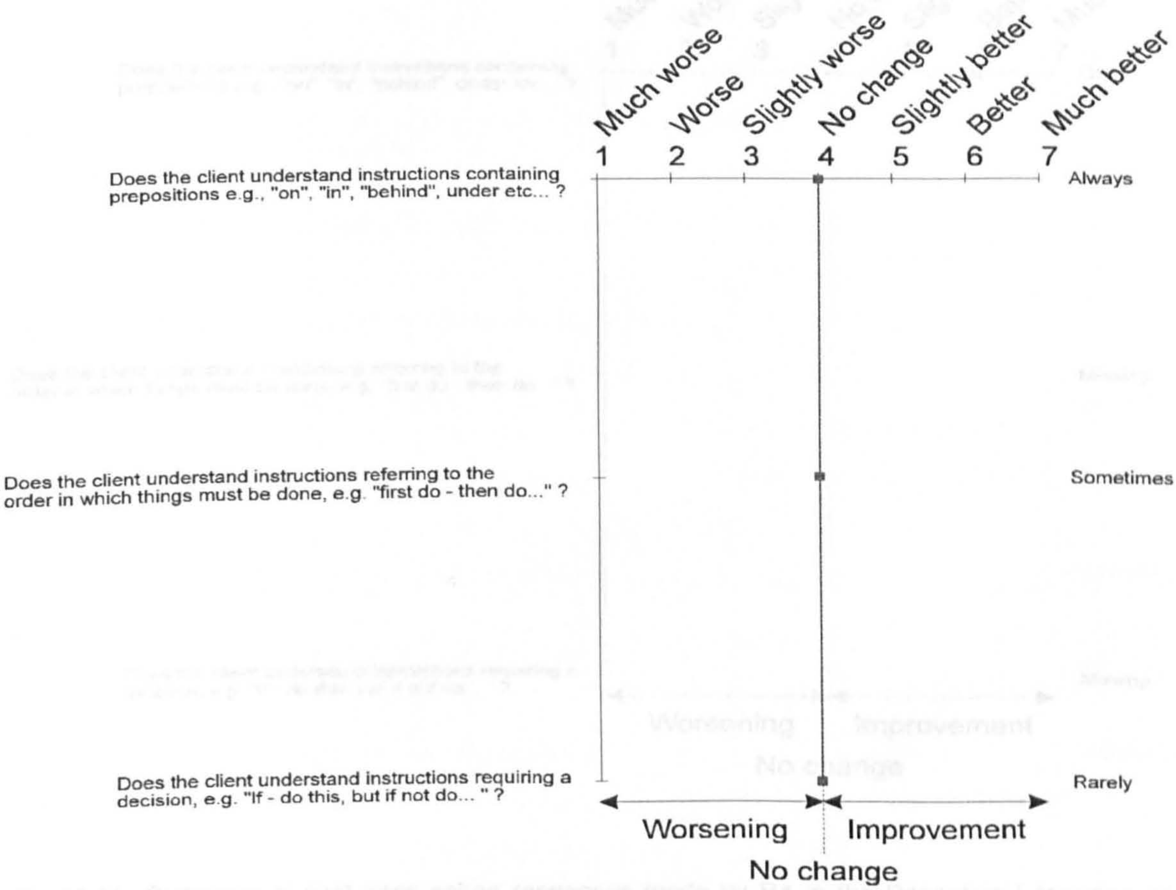


Fig 10.63, Summary of post-intervention responses made by R3 in the Receptive Language domain for U3

Mean	SD	Items responded to by rater
4.00	0.00	3

Fig 10.64, R4 ESPM rating for Receptive Language for U3 (pre- and post-intervention)

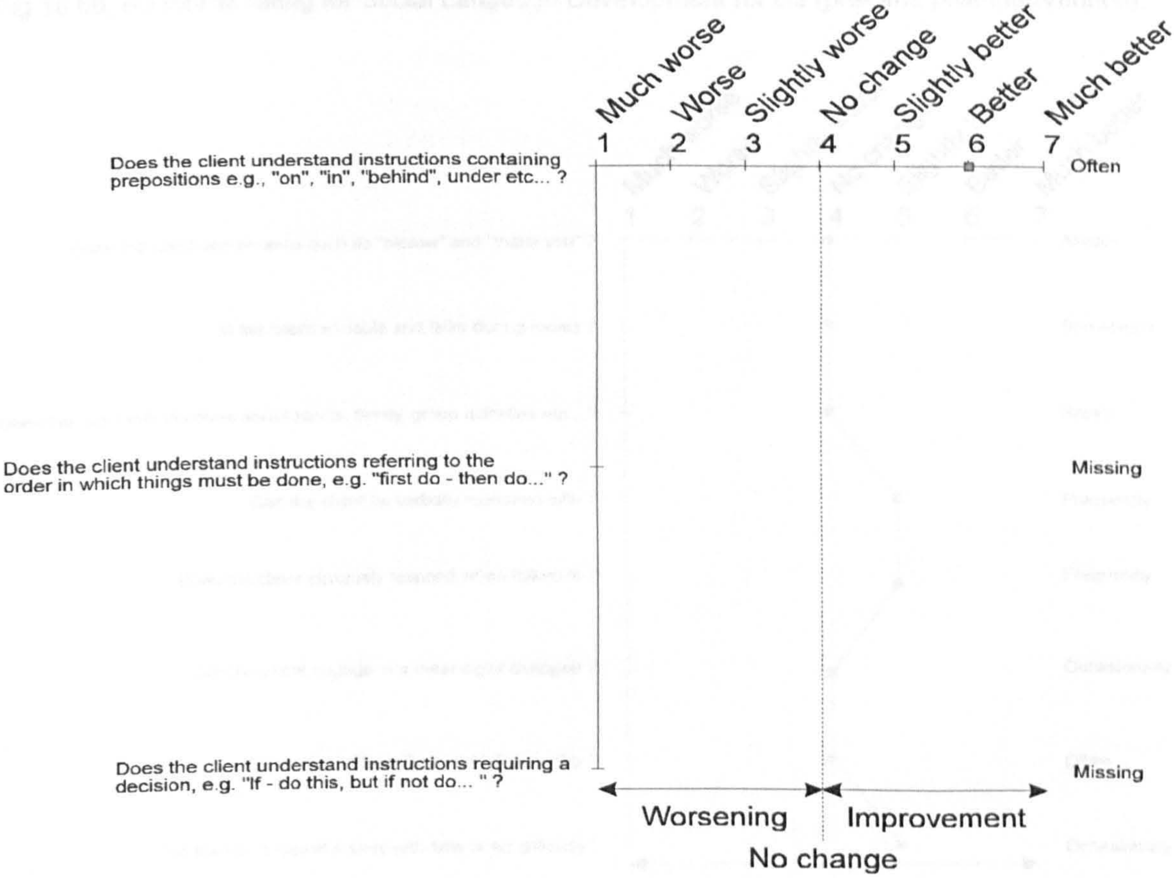


Fig 10.65, Summary of post-intervention responses made by R4 in the Receptive Language domain for U3

Mean	SD	Items responded to by rater
6.00	0.00	1

SOCIAL LANGUAGE DEVELOPMENT DOMAIN RESULTS FOR U3

Fig 10.66, R3 ESPM rating for Social Language Development for U3 (pre- and post-intervention)

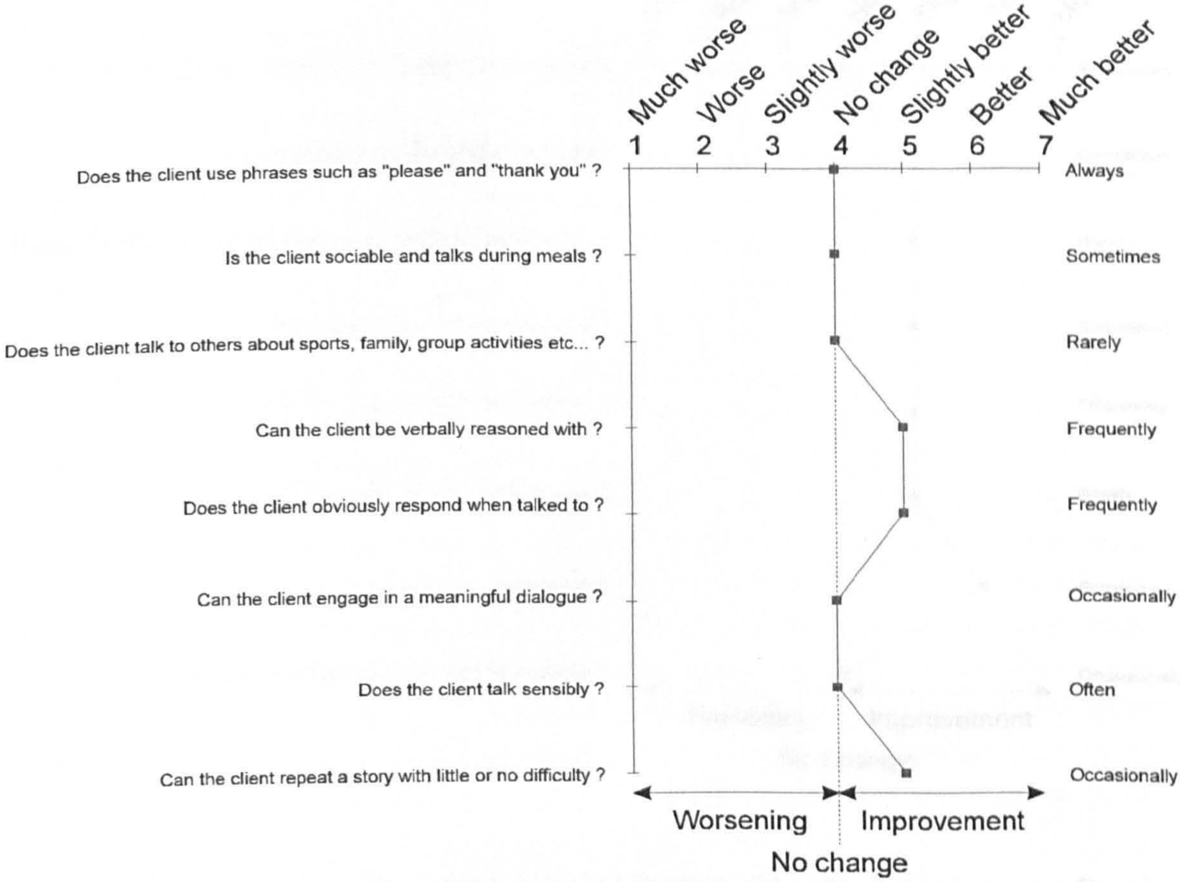


Fig 10.67, Summary of post-intervention responses made by R3 in the Social Language Development domain for U3

Mean	SD	Items responded to by rater
4.38	0.52	8

Fig 10.68, R4 ESPM rating for Social Language Development for U3 (pre- and post-intervention)

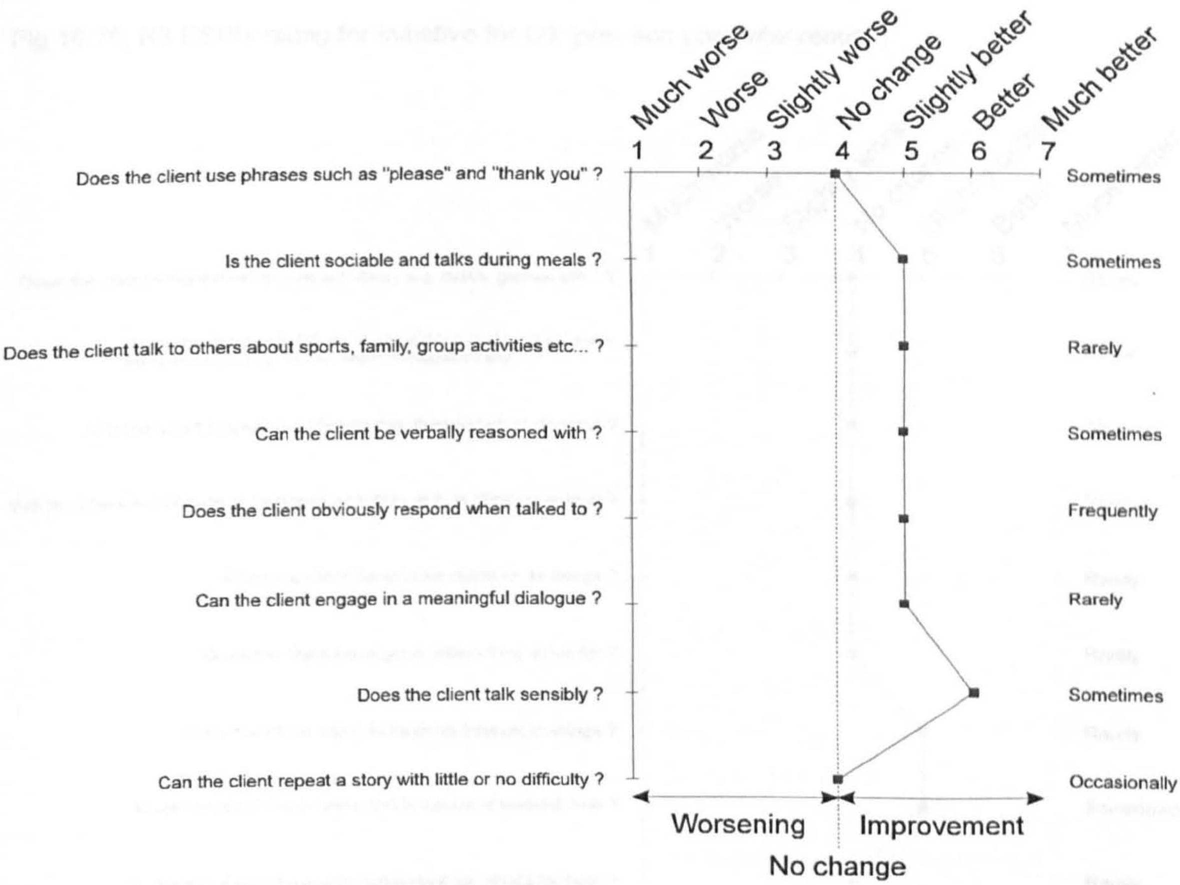


Fig 10.69, Summary of post-intervention responses made by R4 in the Social Language Development domain for U3

Mean	SD	Items responded to by rater
4.88	0.64	8

INITIATIVE DOMAIN RESULTS FOR U3

Fig 10.70, R3 ESPM rating for Initiative for U3 (pre- and post-intervention)

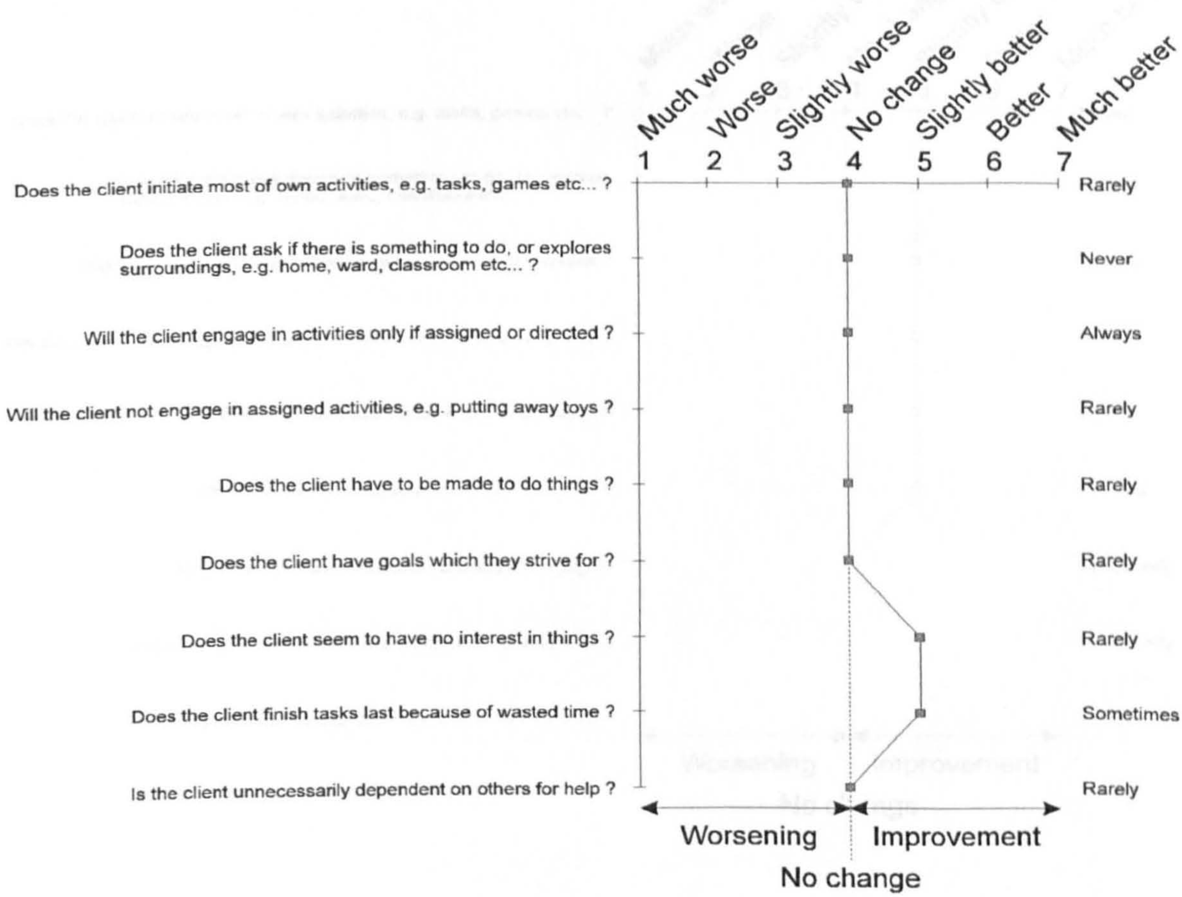


Fig 10.71, Summary of post-intervention responses made by R3 in the Initiative domain for U3

Mean	SD	Items responded to by rater
4.22	0.44	9

Fig 10.72, R4 ESPM rating for Initiative for U3 (pre- and post-intervention)

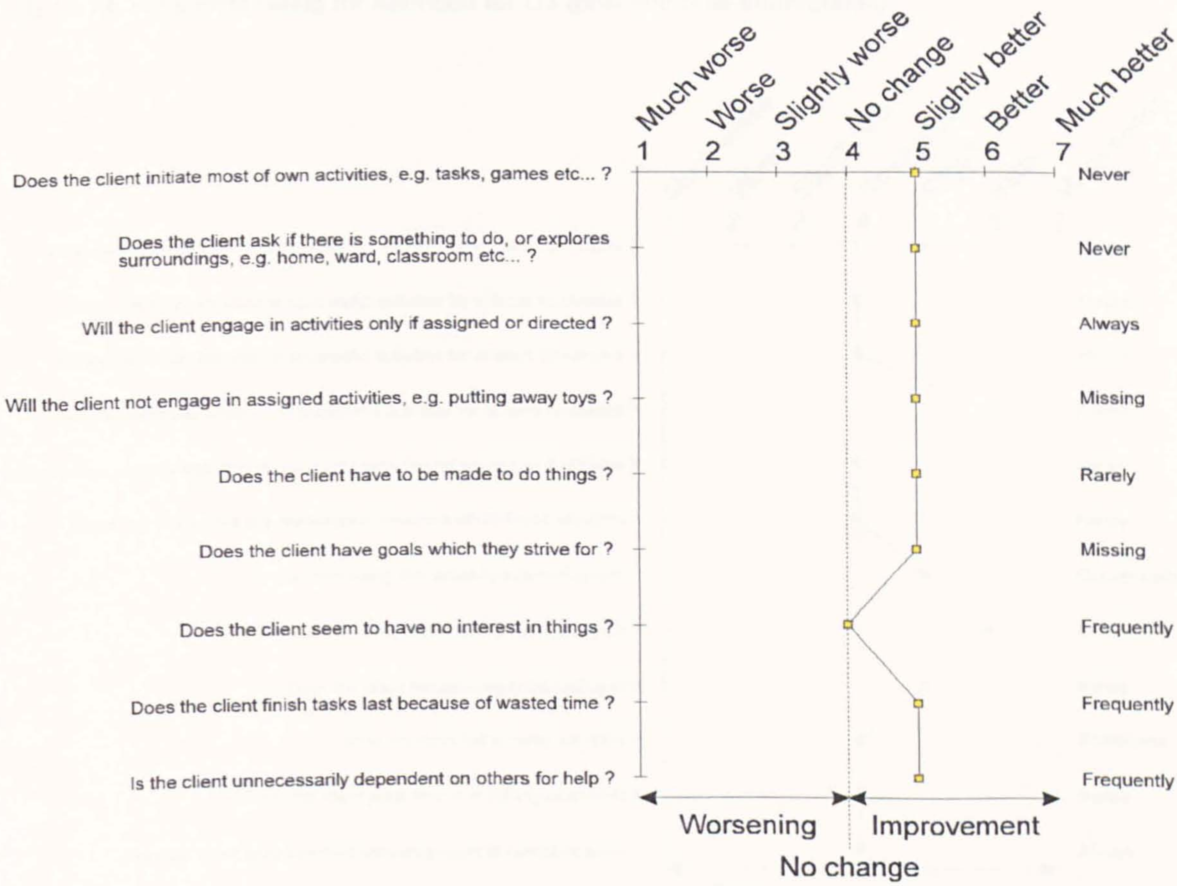


Fig 10.73, Summary of post-intervention responses made by R4 in the Initiative domain for U3

Mean	SD	Items responded to by rater
4.71	0.49	7

ATTENTION DOMAIN RESULTS FOR U3

Fig 10.74, R3 ESPM rating for Attention for U3 (pre- and post-intervention)

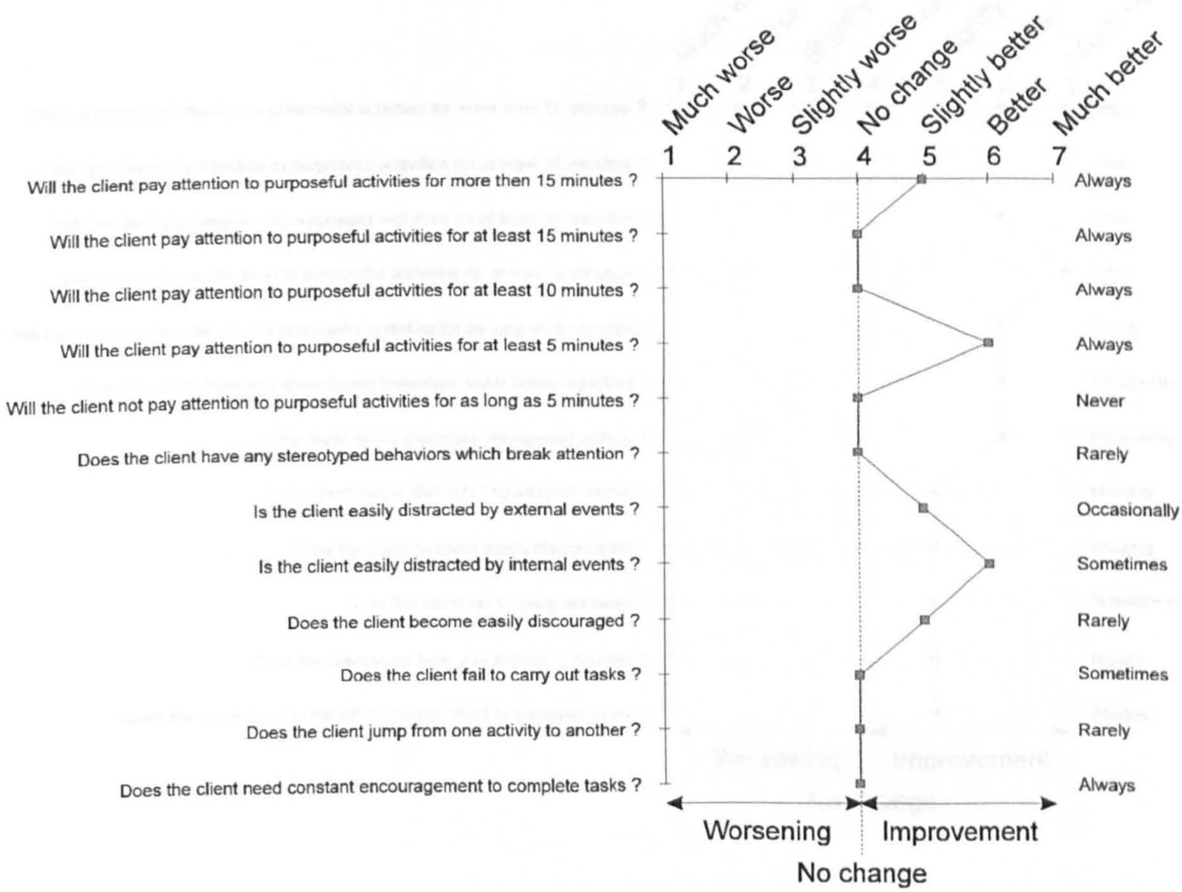


Fig 10.75, Summary of post-intervention responses made by R3 in the Attention domain for U3

Mean	SD	Items responded to by rater
4.58	0.79	12

Fig 10.76, R4 ESPM rating for Attention for U3 (pre- and post-intervention)

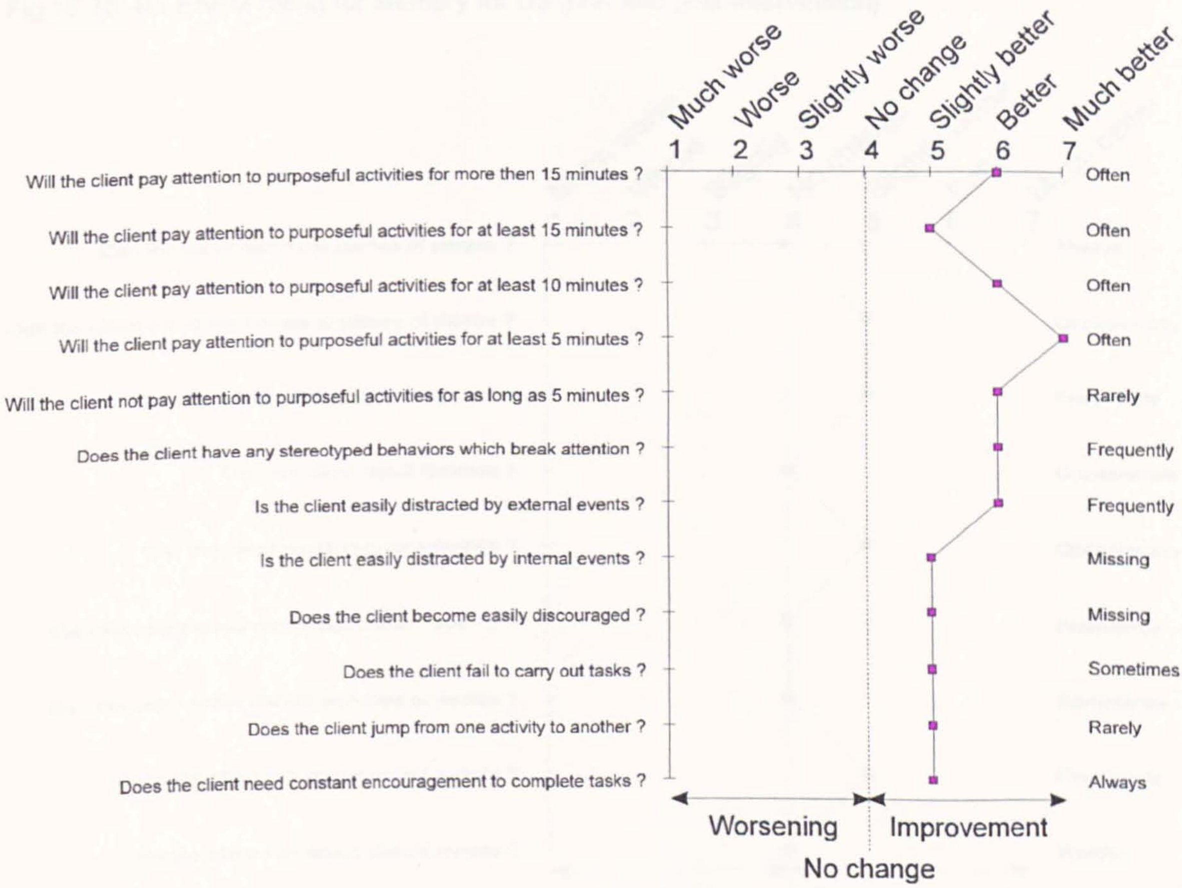


Fig 10.77, Summary of post-intervention responses made by R4 in the Attention domain for U3

Mean	SD	Items responded to by rater
4.40	0.70	10

MEMORY DOMAIN RESULTS FOR U3

Fig 10.78, R3 ESPM rating for Memory for U3 (pre- and post-intervention)

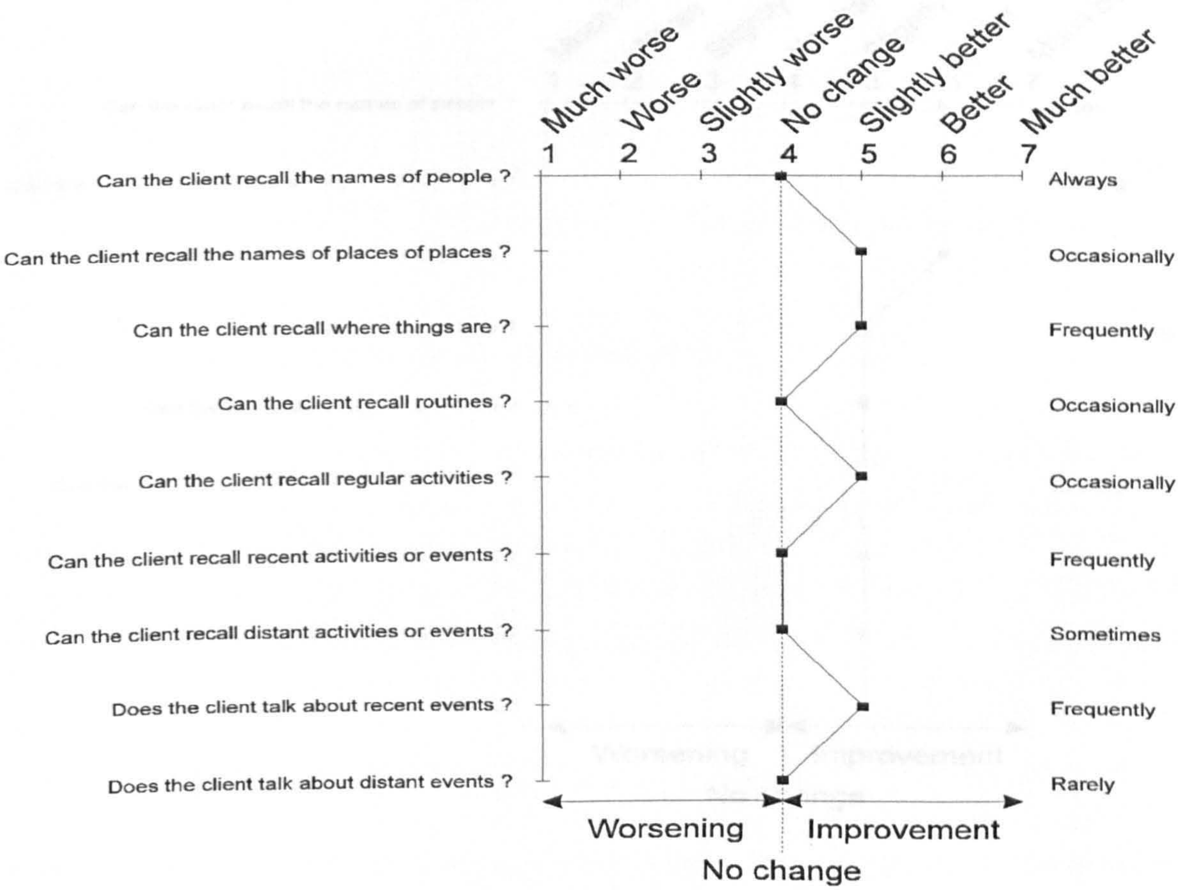


Fig 10.79, Summary of post-intervention responses made by R3 in the Memory domain for U3

Mean	SD	Items responded to by rater
4.44	0.53	9

Fig 10.80, R4 rating for Memory for U3 (pre- and post-intervention)

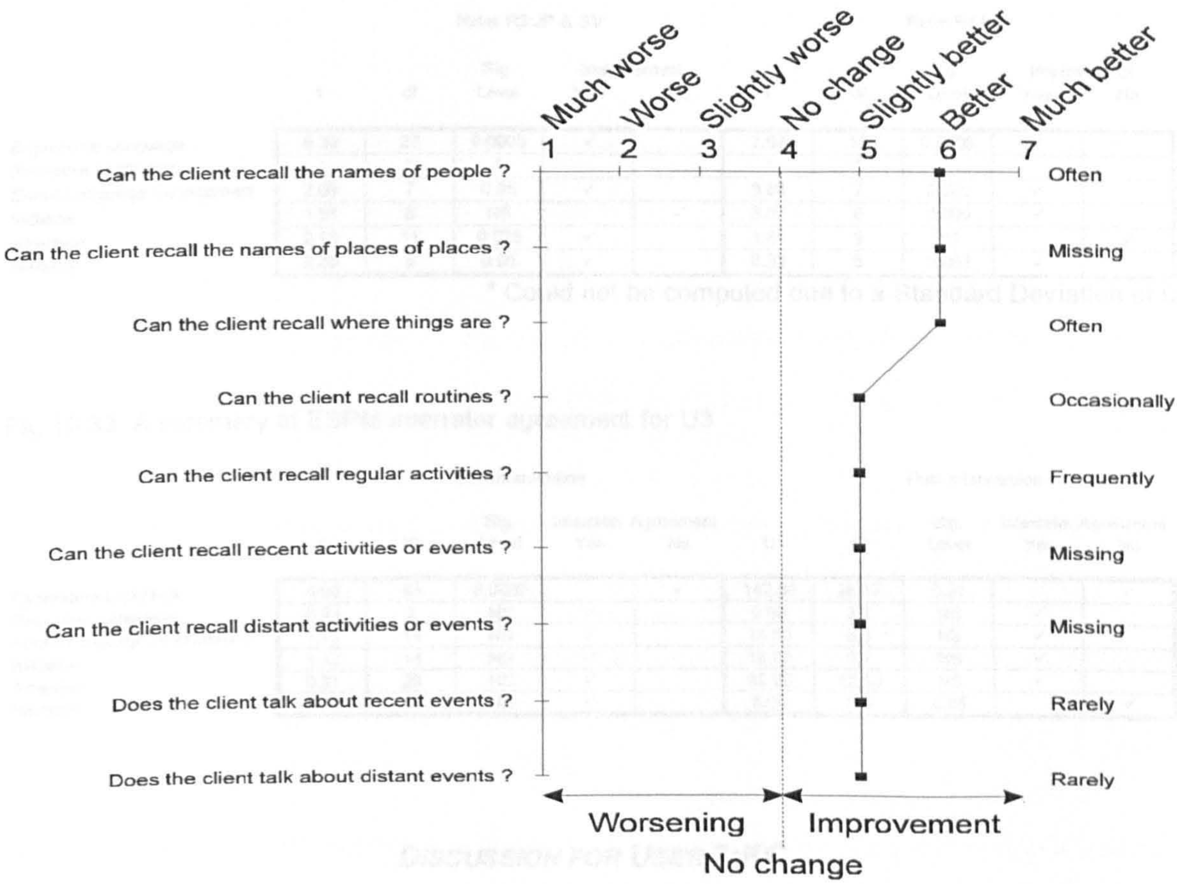


Fig 10.81, Summary of post-intervention responses made by R4 in the Memory domain for U3

Mean	SD	Items responded to by rater
5.33	0.52	6

Fig 10.82, A summary of ESPM post-intervention ratings for U3

	Rater R3:JP & SV					Rater R4:KS				
	t	df	Sig. Level	Improvement Yes	Improvement No	t	df	Sig. Level	Improvement Yes	Improvement No
Expressive Language	6.30	27	0.0005	✓		7.67	16	0.0005	✓	
Receptive Language	*	*	*			*	*	*		
Social Language Development	2.05	7	0.05	✓		3.86	7	0.005	✓	
Initiative	1.51	8	NS		✓	3.87	6	0.005	✓	
Attention	2.55	11	0.025	✓		1.81	9	NS		✓
Memory	2.53	8	0.05	✓		6.32	5	0.001	✓	

* Could not be computed due to a Standard Deviation of 0

Fig 10.83, A summary of ESPM interrater agreement for U3

	Pre Intervention					Post Intervention				
	t	df	Sig. Level	Interrater Agreement Yes	Interrater Agreement No	U	df	Sig. Level	Interrater Agreement Yes	Interrater Agreement No
Expressive Language	4.60	43	0.0005		✓	142.00	28,17	0.01		✓
Receptive Language	0.23	2	NS	✓		0.00	3,1	NS	✓	
Social Language Development	1.13	14	NS	✓		18.50	8,8	NS	✓	
Initiative	1.32	14	NS	✓		16.00	9,7	NS	✓	
Attention	0.31	20	NS	✓		52.50	12,10	NS	✓	
Memory	0.65	13	NS	✓		8.00	9,6	0.05		✓

DISCUSSION FOR USER 3:KC

Both raters agreed U3 had improved significantly in three domains. Namely, Expressive language, Social language development and memory. R3 also rated an improvement in attention, whereas R4 did not. R4 rated an improvement in initiative, whereas R3 did not.

Within the Expressive language domain both raters rated an improvement in U3's vocal level (loudness) over it's previous faintness. Raters agreed that her speech was more fluent as a result of reduced tendency toward dysphasia and reduced likelihood of the speech flow being disrupted by other irregular interruptions. Combined with ratings of reduced dysatharia speech was considered many times more intelligible. These marked improvements may account for the ratings of lessened tendency to be rambling/incoherent as it may simply be that prior to intervention she was less likely to be understood. Further support for this assertion comes from ratings which suggest U3 was now more likely to be understood by both those who know her as well as by those who are unacquainted. The degree to which U3 asked "why" questions was also rated as showing improvement. U3's general ability to correctly name and describe people and objects in pictures rated as having improved.

In terms of Social language development both raters agreed that U3 was more apt to be verbally reasoned with that she had been previously. Raters also rated a noticeable improvement in terms of her obviously responding when talked to. Previously she had been quite variable in her response to direct linguistic interaction. R4 also rated improvements in U3's general language sociability, e.g. whether she talked to others about sport, family and other group activities.

Rater R4 rated improvements on all items within the memory domain suggesting a real gain. Both raters agreed that she was now more likely to remember the names of people and places than before. In general they rated how she was also more likely to remember where things are, e.g. where she had left her handbag. Her short term recall was rated as having improved notably also.

Raters could not agree on significant improvements in two areas, namely Initiative and Attention. Despite lack of interrater agreement R4 did rate significant improvements within these areas. Within the Initiative domain R4 rated improvements in all but one of the nine items. R4 rated U3 as being more likely to initiate most of her own activities. She was also rated as being generally more independent e.g. engaging in activities only when told to do so and being more likely to ask for things to do when idle. In motivation terms R4 noted how U3 was more likely to set both interim and long term goals for which she would strive. As a corollary, motivation levels were noted as having increased. In addition to being more highly motivated U3 was also rated as being more independent in general. She was rated as being less dependent on others for help and as being less likely to finish tasks last due to wasted time or inefficiency.

Within the Attention domain R3 rated improvement in both short term and long term attention spans. Five and fifteen minutes respectively. U3's distractibility was also rated as having reduced. This reduction was noted as applying to both internal and external distracting events. For example, distracting thoughts were less likely to occur as on-task attention increased; and noises from outside the room were also less distracting. Reduced distractibility can also be thought of as contributing to the rating that U3 was less likely to become easily discouraged.

RESULTS FOR USER 4: CP

From its inception the ESPM questionnaire was designed to have each user rated by two raters who had known them for at least a year. However in the case of U4 this was not possible, as the second potential rater had left to take up another post immediately prior to intervention. Despite this shortcoming it was decided that U4 should take part in the study as potentially intervention could be highly beneficial. Obviously interrater agreement could not be evaluated, and therefore analysis concentrated on R4's ratings alone; which are considered to be none-the-less valid as a result.

Fig 10.84, Summary of raters relationship to U4

Rater	Relationship to user	How long known user (years)	Ave weekly contact (hours)
R4*	Head of Dept	1	2

* R4 is the same rater who rated U3

EXPRESSIVE LANGUAGE DOMAIN RESULTS FOR U4

Fig 10.85, R4 ESPM rating for Expressive Language for U4 (pre- and post-intervention)

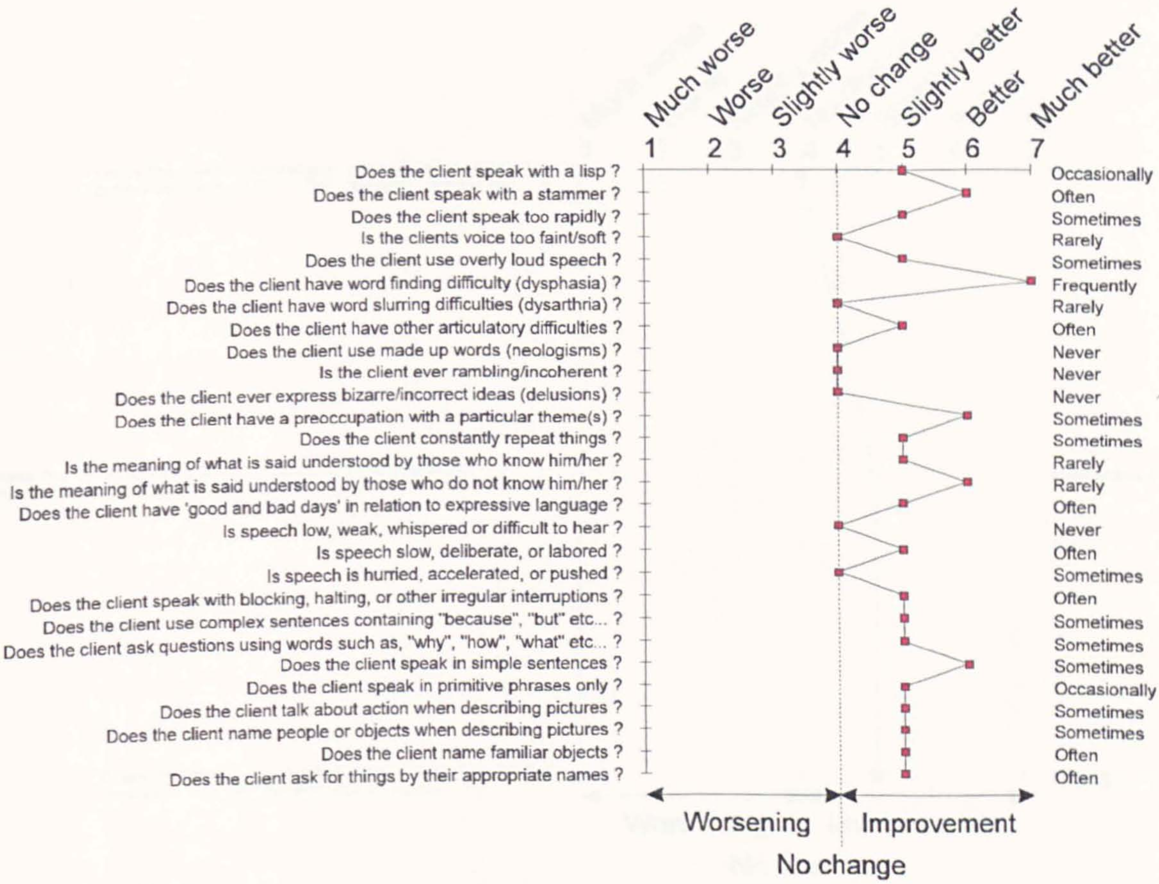


Fig 10.86, Summary of post-intervention responses made by R4 in the Expressive Language domain

Mean	SD	Items responded to by rater
4.96	0.74	28

RECEPTIVE LANGUAGE DOMAIN RESULTS FOR U4

Fig 10.87, R4 ESPM rating for Receptive Language for U4 (pre- and post-intervention)

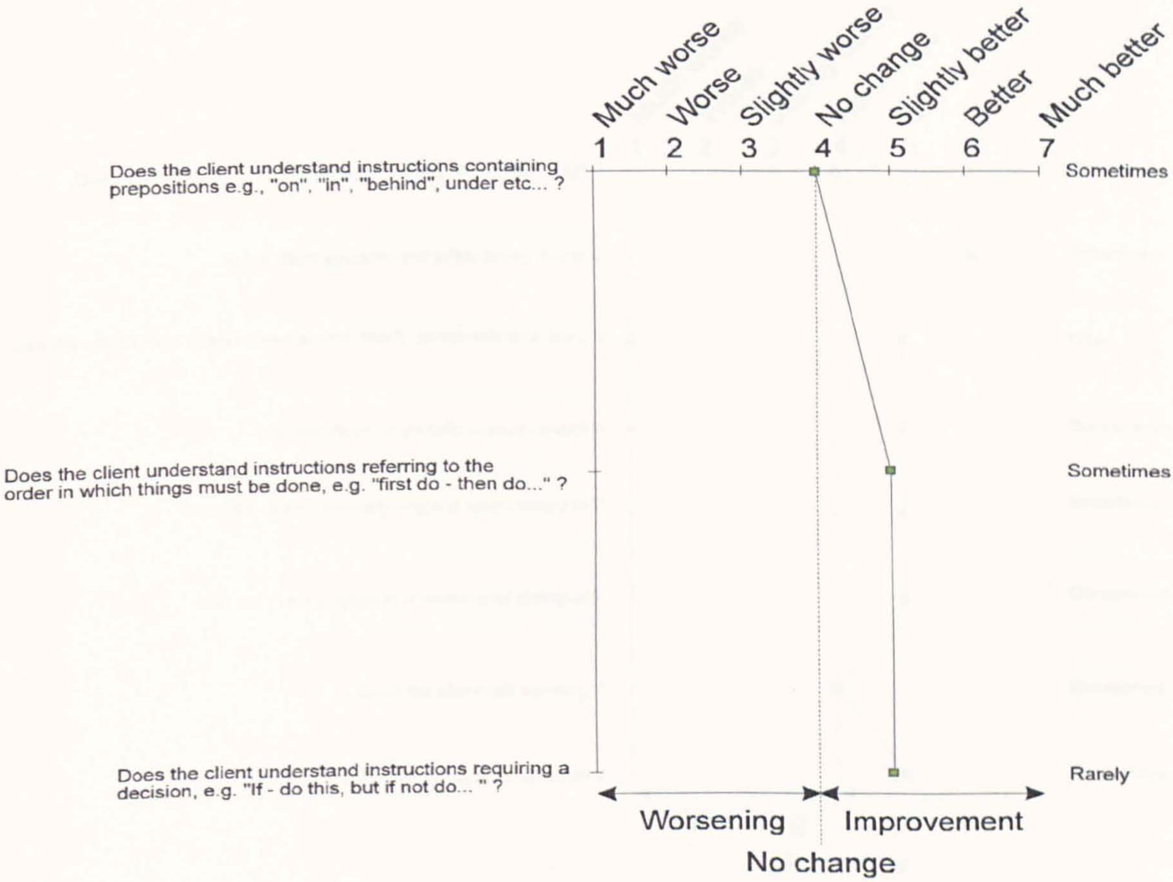


Fig 10.88, Summary of post-intervention responses made by R4 in the Receptive Language domain for U4

Mean	SD	Items responded to by rater
4.66	0.58	3

SOCIAL LANGUAGE DEVELOPMENT DOMAIN RESULTS FOR U4

Fig 10.89, R4 ESPM rating for Social Language Development for U4 (pre- and post-intervention)

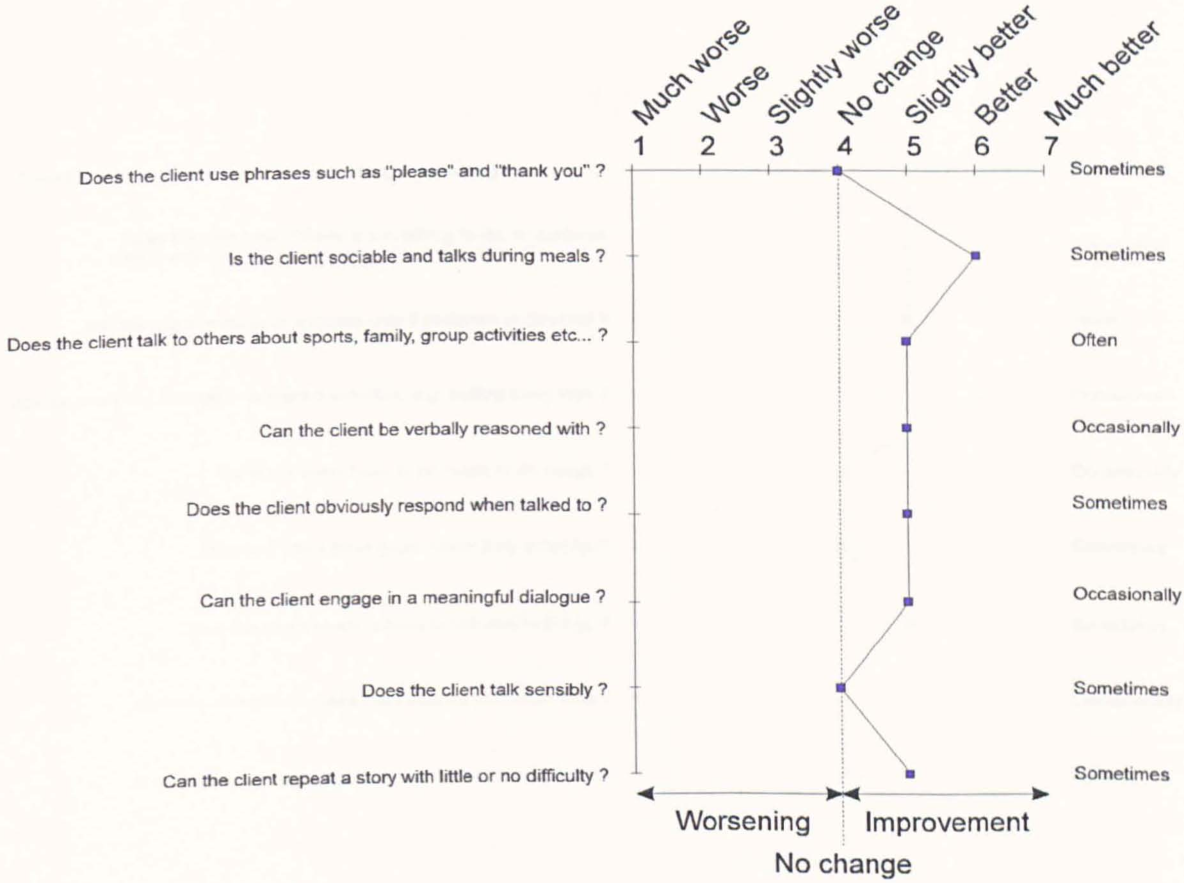


Fig 10.90, Summary of post-intervention responses made by R4 in the Social Language Development domain for U4

Mean	SD	Items responded to by rater
4.88	0.64	8

INITIATIVE DOMAIN RESULTS FOR U4

Fig 10.91, R4 ESPM rating for Initiative for U4 (pre- and post-intervention)

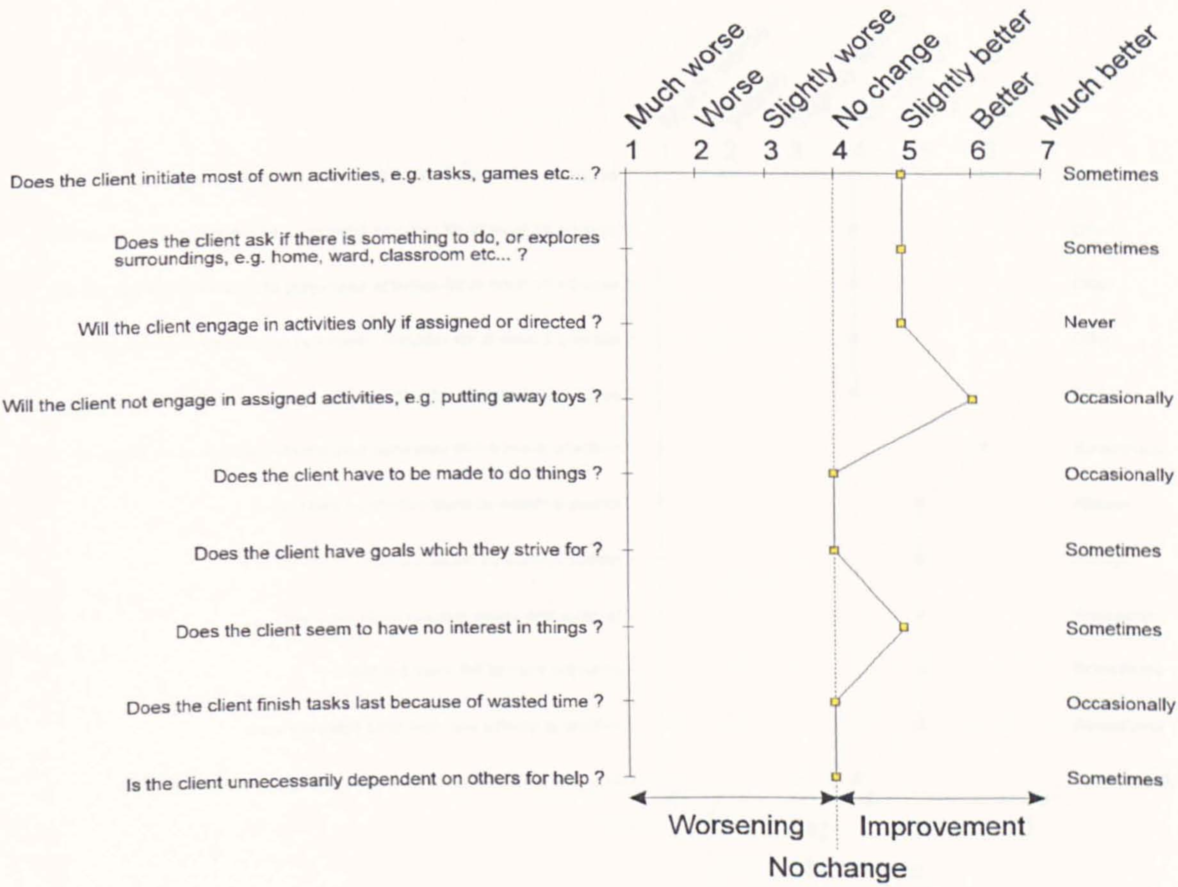


Fig 10.92, Summary of post-intervention responses made by R4 in the Initiative domain

Mean	SD	Items responded to by rater
4.66	0.71	9

ATTENTION DOMAIN RESULTS FOR U4

Fig 10.93, R4 ESPM rating for Attention for U4 (pre- and post-intervention)

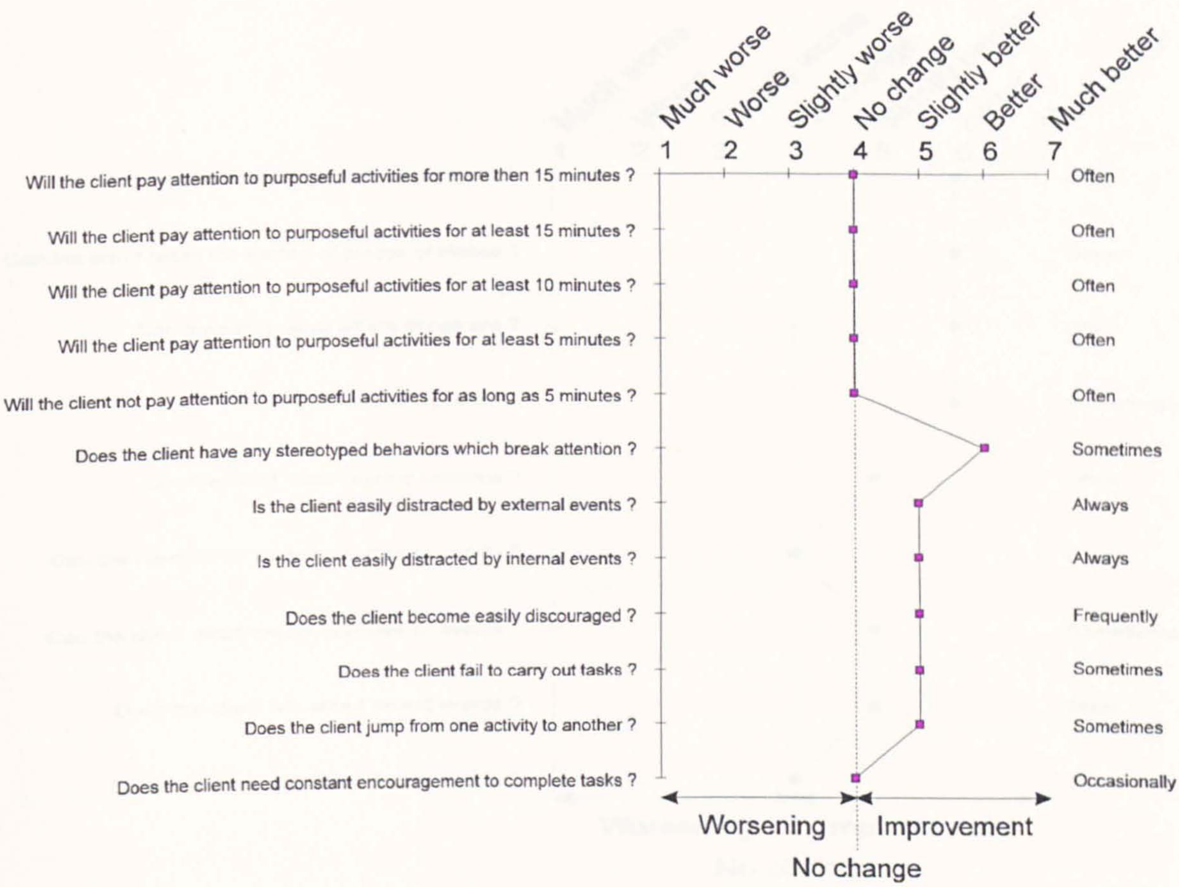


Fig 10.94, Summary of post-intervention responses made by R4 in the Attention domain for U4

Mean	SD	Items responded to by rater
4.58	0.67	12

MEMORY DOMAIN RESULTS FOR U4

Fig 10.95, R4 ESPM rating for Memory for U4 (pre- and post-intervention)

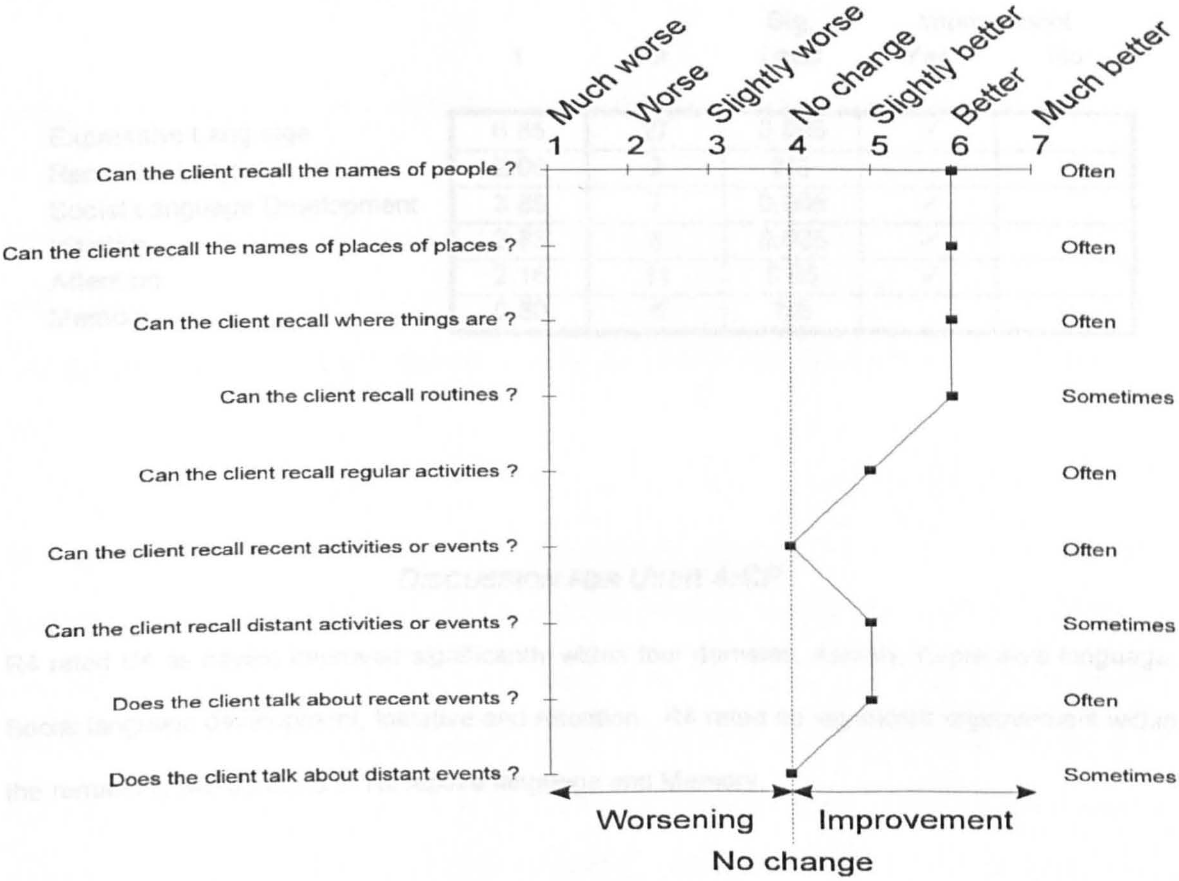


Fig 10.96, Summary of post-intervention responses made by R4 in the Memory domain for U4

Mean	SD	Items responded to by rater
5.22	0.83	9

Fig 10.97, A summary of ESPM post-intervention ratings for U4

Rater R4:KS

	t	df	Sig. Level	Improvement Yes	Improvement No
Expressive Language	6.85	27	0.005	✓	
Receptive Language	2.00	2	NS		✓
Social Language Development	3.86	7	0.005	✓	
Initiative	2.83	8	0.025	✓	
Attention	2.16	11	0.05	✓	
Memory	0.80	8	NS		✓

DISCUSSION FOR USER 4:CP

R4 rated U4 as having improved significantly within four domains, namely, Expressive language, Social language development, Initiative and Attention. R4 rated no significant improvement within the remaining two domains of Receptive language and Memory.

Within the Expressive language domain R4 rated U4's pronounced stammer as being less apparent than it had been previously. Her rapid speech rate and slight lisp were also rated as having improved. Crucially her frequent dysphasia was rated as showing a very significant reduction. Her tendency toward preoccupation with a particular theme was also rated as having reduced. It is possible that this apparent preoccupation may have been due to her not being understood and therefore repeating utterances constantly. This tends to be confirmed by the author, who understood her speech to a good degree and tended to notice less repetition than say other unfamiliar staff. Possibly as a result R4 rated U4 as being more likely to be understood by those who did not know her after intervention. Lastly the complexity of utterances were also rated as becoming more complex and to a degree normalised.

In terms of Social language development U4 was rated as becoming more sociable during meals and other social gatherings, and was rated as being generally more conversant. U4 was more likely to respond to when talked to and was more apt to be verbally reasoned with than she was previously disposed to be. She was also rated as being more likely to engage in meaningful dialogues and was also better at repeating a story she had been told.

Initiative also showed improvement with ratings suggesting that U4 was more likely to initiate most of her own activities now. Furthermore, if there was nothing to do she would now be more likely to ask for a task to complete or would engage on a task of her own volition. U4 was also rated as being more likely to engage is activities she was directed to complete. In sum she was rated as having more interest in things in general and can be thought of as having higher levels of motivation.

Disappointingly attention span remained largely unchanged from preintervention levels. However the stereotyped behaviours she displayed prior to intervention were less marked and were accordingly rated as being less likely to break her short attention span. She was also rated as being less distracted by both internal and external events. Rater R4 rated her as being less apt to become discouraged and as a result less likely to fail to carry out tasks. Finally she was rated as being less likely to jump from one activity to another. Although her actual attention span was not thought of as having increased markedly, this was compensated for to a large degree by her increased levels of on-task focusing.

GENERAL RESULTS

All users were rated by at least one rater as having improved within three or more domains. More importantly no user worsened on any domain. Obviously statistical comparisons between users are meaningless given users' differing abilities and handicaps. However, it is useful to compact responses for all six domains for each user by rater and then carry out a single-sample t-test in a similar way to analyses carried out earlier. This will test the hypothesis that post-intervention ratings are significantly different from baseline when considered over all domains. Given the direction of the difference one can say whether this difference represents a significant improvement or worsening overall over the expected "no change" over baseline.

Fig 10.97, A summary of post-intervention ratings over all six domains by user and rater

	First Rater						Second Rater					
			Sig.	Improvement					Sig.	Improvement		
	t	df	Level	Yes	No	Rater	t	df	Level	Yes	No	Rater
U1	4.55	66	0.0005	✓		R1	11.02	68	0.0005	✓		R2
U2	2.54	68	0.01	✓		R1	13.27	68	0.0005	✓		R2
U3	7.15	68	0.0005	✓		R3	9.52	48	0.0005	✓		R4
U4	9.95	68	0.0005	✓		R4	*	*	*	*	*	*

* Could not be computed as a second rater was unavailable for U4

Analysis confirmed that after six months (24 weeks) of Easy Speaker intervention all users had significantly improved when all behavioural domain measures of language and cognition were combined. Further support is given for real improvements in users as measured behaviourally as all four raters ratings suggested a highly significant improvement overall.

GENERAL DISCUSSION

Given users various handicaps and abilities it is encouraging that all users were rated by at least one rater as having improved significantly within three or more domains, and that no user worsened in any domain, or more precisely, on any of the ESPM questionnaire's 68 items. One would suggest that this lends a fairly good degree of external validity to the hypothesis that improvements shown in the earlier skill acquisition study, discussed in chapter nine, are mirrored in off-device improvements in both language and cognition when measured in terms of adaptive behaviour. Support is given to earlier informal studies of Easy Speaker intervention and off-device improvements, e.g. Rostron, Plant and Hermann (1994). Indeed it is this notion of off-device improvements in language and cognition that is central to the ethos of Easy Speaker intervention for the more able of the Learning Disabled population. It is hoped that less able non-vocal users would be able to use Easy Speaker as a primary communication channel and in addition benefit from secondary improvements in language and cognition shown by their more able peers. The most heartening aspect of such improvements is that the majority of users improved in domains that they were previously poor performers in, e.g. in the case of U1, his level of attention was enhanced markedly where previously it had been relatively short.

Despite each user being unique and to all intents and purposes treated as a single subject, it is useful to attempt to identify areas where Easy Speaker intervention appeared to have consistent outcomes behaviourally. One domain in which all users were rated as having significantly improved in by at least one rater is Expressive language. This is one of the central domains where Easy Speaker intervention was targeted for more able users. In addition to the obvious cognitive deficits all users displayed prior to intervention, marked vocal and linguistic impairments affected the intelligibility of their speech both to known others and strangers. Expressive language then encompassed not just aspects of quality, but also consistency, complexity, and semantics. Improvements were noted in all areas.

Whilst involved in Easy Speaker intervention users were encouraged to mimic digitised speech as it was output, and then as sessions progressed to pre-empt it where possible. All users at some stage recorded their own residual speech and associated it with specific pictorial representations, or symbols, with the aid of the author or other members of staff. Although Easy Speaker was not intended to act as a speech therapy replacement, it was designed to enhance or argument the users own speech where applicable. It might be considered to be a more global communication therapy tool however. Three of the four users had when younger received various forms of formal speech therapy. However none had received any such therapy within the last five to ten years or so. This was either due to their static progress, resource demands or for other reasons, e.g. moving away from the area.

One might propose that such improvements in Expressive language were due to a combination of factors inherent in Easy Speaker usage. That is, high quality speech output, the ability to receive a form of self-paced covert speech therapy, and the highly motivating and instant reward basis of the system itself. It is well known for example that instant gratification of a high quality nature is likely to act as a strong reinforcer, and that if this is initiated by the user themselves the stronger this is likely to be. Such a system can almost be thought of as giving 'communication therapy by osmosis'. The key issue being one of motivation leading to imitation, and through shaping via instant gratification, to a closer approximation of correct phonetic patterns and other related linguistic skills. In less able non-vocal users such reinforcement and gratification is seen as crucial for learning to use the system.

One might go on to suggest that formal speech and communication therapy for a more able user group might be better augmented through a system such as Easy Speaker. One can imagine the difficulties in administering formal speech and communication therapy to an individual with a moderate to severe Learning Disability, with short attention span, low motivation, and depressed self-esteem.

Another domain in which all users were rated as having significantly improved in by at least one rater is Social language development. Improvements in this domain are closely related to improvements in Expressive language. The difference being that Expressive language deals with aspects of output and content, whereas Social language development is concerned with to whom and in what manner the user interacts linguistically. For example, the question of whether the "client is sociable and talks during meals" is addressed, as is whether the "client can engage in a meaningful dialogue" etc... Hopefully as the users Expressive language improves through Easy Speaker interaction one would hope to see an increase in Social language interactions, whether this be on or off-device. In this more able user group this proved to be the case, as highlighted by raters, and more subjectively, by the comments of other members of staff not directly involved in the study. This could be due to a combination of the user being more likely to be understood by others, and the user themselves realising their own residual speech had improved slightly and being more likely to attempt to interact linguistically with others in social settings. Obviously increased social interaction of a successful nature is a crucial component in further enhancing a users linguistic ability and can be thought of being itself a reinforcing agent. Undoubtedly quality of life will also be enhanced as individuals would be less likely to remain socially passive as before.

Two other domains in which all users were rated as having significantly improved in by at least one rater are Initiative and Attention. Although semantically different it is useful to consider these two domains as being complementary. That is, the initiative domain encompasses motivation, and where increased motivation is induced, positive affects on attention are the likely result.

Initiative was concerned with such aspects as whether the "client initiated most of their own activities", whether the "client has to be made to do things", and whether the "client asks if there is something to do" etc... One can reduce these behaviours down to self-directed behaviour and to goal setting. Logically increased motivation would lead to increased initiative, and hence increases in self-directed behaviour and goal setting. As with the domains discussed previously this would be likely to have a knock-on effect as an intrinsic reinforcer. That is, a user may increasingly interact with their environment due to reinforcers created as a result of their own initiatives. One would hope this might become a self-feeding loop, where the user would provide their own intrinsic motivational rewards. Such increases in initiative and motivation were seen in users both during Easy Speaker usage and to a lesser, but noticeable extent, off-device. Within the initiative domain on average raters noted an improvement on five of the nine items, with some raters noting improvements on up to eight for individual users.

Attention was concerned with such aspects as how long the user would pay attention for, whether the "client was easily distracted", and whether the "client jumps from one activity to another" etc... Whilst using Easy Speaker the users previously measured, or for that case estimated, attention span seemed unimportant and unrelated to how long the user spent interacting with Easy Speaker. The crucial factor instead appeared to be increased levels of motivation and reward. These phenomenal increases in on-device attention were observed to transfer through to off-device situations, although as with initiative and motivation, to a lesser extent. Lesser in this case meaning, marked and easily observable as compared with their functioning prior to intervention, but not to the same highly elevated levels as when engaged on-device. One can summarise these improvements within the attention domain by stating that on average users improved on seven out of twelve items, with some raters noting an improvement on all items. This suggests a good degree of increased attention transfer to off-device situations.

The final domain in which users showed off-task improvements is memory. Memory was concerned with such aspects as, "can the client recall where things are", "can the client recall distant activities or events", and "can the client recall regular activities" etc... Again to be able to utilise Easy Speaker fully each user had to bring their limited cognitive abilities into play. This was not formally empathised, but was a function of general Easy Speaker usage. Indeed without mapping the Easy Speaker concept hierarchy on to their own, they would have been unable to use the system it at all. Although cognitive load was designed to be lower than other symbolic non-hierarchical systems, a certain degree of rote memorisation was involved. This ability, to memorise the hierarchy was demonstrated in the skill acquisition study discussed in the previous chapter. Indeed such patterns of skill acquisition demonstrate learning has occurred. It is this increased fertility for learning and application of memory off-device that was measured behaviourally within the memory domain. One can summarise these improvements by stating that on average users improved on five out of nine items, with some raters noting improvements on all items.

One domain which in which users showed little significant improvement is that of receptive language. This can be attributed to a statistical artefact due to the small number of questionnaire items which made up the domain, namely three. Unfortunately the ABS and FPR only contained three exclusive items regarding receptive language despite offering many more for other domains with each having over 150 and 300 assorted items respectively. In the interest of consistency it was decided no to supplant the original items in order to increase the overall number within the domain. It was assumed that with raters already being familiar with the ABS and FPR that any additions may affect reliability and validity, especially if the items were untested clinically. Despite users failing to achieve statistically significant progress, all did improve behaviourally. Such improvement is reflected in many of the other questionnaire items which have a receptive language component, e.g. "Talks socially at meal times".

Although both the Adaptive Behaviour Scales and Functional Performance Record are well accepted clinical inventories for measuring adaptive behaviour of clients in the real world, drawing inferences from them can be difficult. For example, they attempt to "measure" memory not by strict experimentation, but through the client's behaviour in their environment, e.g. "Can the client recall the names of people ?". In clinical settings, behavioural improvement, although a cruder measure, is more relevant both to the client themselves and the practitioner. Although many of the improvements rated in Easy Speaker users were clear behaviourally, measuring them precisely would have been impossible. For example, a memory span test pre- and post-intervention would have been unlikely show any change, whereas a behavioural measure would. Each behavioural measure then can represent a culmination of factors that make up an improvement, e.g. improvements in attention and motivation can affect memory.

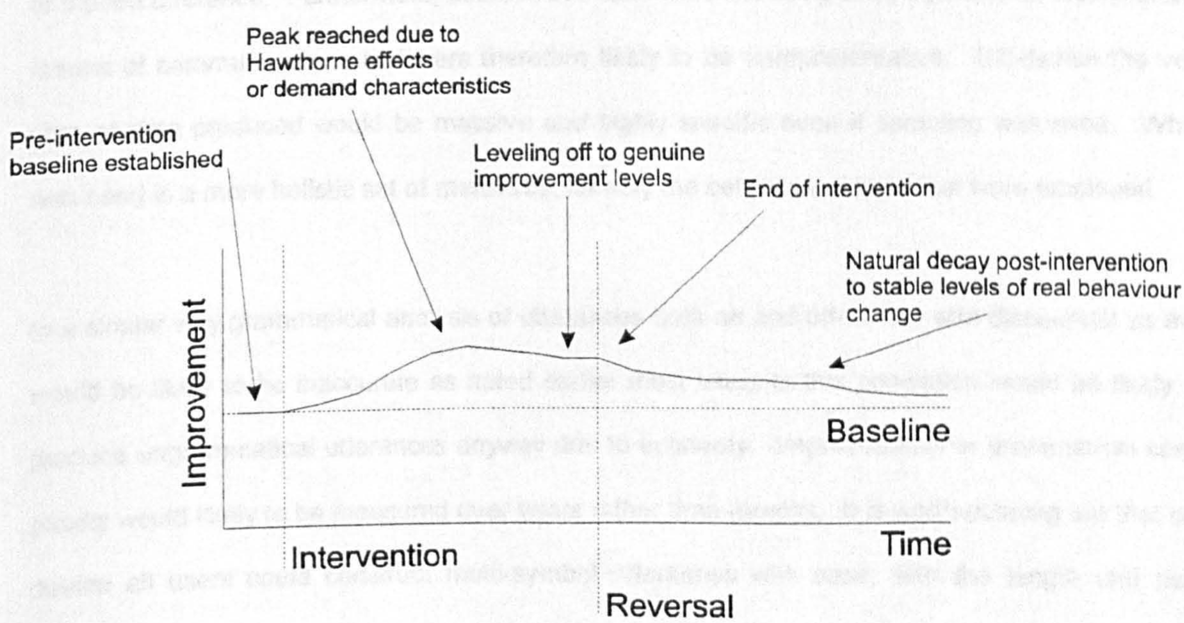
Such clinical inventories, and indeed the ESPM questionnaire, can on the surface appear to be speculative due to the fact that ratings are to a large extent subjective. With average levels of inter-rater agreement this view can be compounded. However, in clinical settings such an inventory is very useful as it concentrates on what the client can do, rather than on what they can't. Domains should be considered indicative and not conclusive.

Average levels of inter-rater agreement does raise the question of demand characteristics in those completing the questionnaires. However, due to the experienced nature of all the respondents this is thought not to be the case, and is probably a reflection of the difficulty in assessing behavioural indicators per se. If the author alone had completed all questionnaires this would have been likely to introduce unintentional experimenter bias compounded by the lack of experience in clinically ratings clients behaviour. Fortunately all raters were already familiar with both the ABS and FPR prior to involvement in the study as they were required to complete them regularly as part of their jobs.

Unlike the pure performance measures taken by ARTS demand characteristics within the users themselves must also be considered. For example, did the presence of the author over the 24 week period affect their behaviour unduly, e.g. did they become more sociable, or "talk during meal times ?". Thus ultimately did this affect questionnaire ratings ? Given the users in question this is unlikely to have been the case for several reasons. Firstly all had been, and were still, receiving input from clinical psychologists and assistants throughout. Therefore one might assume that any effects as a result of their intervention would have already been shown, e.g. Hawthorne, or demand effects leading to increased sociability etc..., and were unlikely to be increased by the intervention of the author. Secondly by the time the post-intervention questionnaire had been administered any such Hawthorne effects would may have worn off as users became more use to the author and Easy Speaker. That is, most improvement in adaptive behaviour would have been likely to have been observed during the first half of the study (weeks 1-12) and any improvements still present post-intervention (week 24) would likely to be more stable and a direct result of Easy Speaker intervention. In general for Hawthorne effects to have a significant impact on data collection there is usually some incentive for the participants to please the experimenter. However, in this case it is probable that the reverse was true, as generally in such environments "innovation" was looked on unfavourably. That is, there was little incentive for raters to be more favourable than the users behaviour warranted.

In order to test the stability of the behavioural improvements suggested by the ESPM questionnaire a reversal design (Elmes, Kantowitz and Roediger 1992) could have been employed. Using such a design one would expect the following patterns of behavioural change to be shown :-

Fig 10.98, Monitoring improvements in adaptive behaviour using a reversal design



Under this design at the end of intervention, or reversal point, the ESPM questionnaire would have been administered, then during the post-intervention period several more questionnaires would be administered to monitor the stability of any behavioural changes. In other words, the decay back toward baseline, or pre-intervention levels. The less decay, the more permanent the behavioural change brought about by Easy Speaker usage.

This was not done in the current study for three reasons. The first being that more Learning Disabled non-vocal users would not be subjected to device withdrawal and so with constant device availability would hopefully continue to display a steady rate of improvement. That is, if the device acted as their primary means of communication. Secondly, by re-administering the same questionnaires to the same caregivers repeatedly over a short period would undoubtedly lead to increased inaccuracies of measurement. Finally, in this particular case half of the sample were residents in a long term care facility that was due to close after the 24 week study period was complete making follow-up difficult, if not impossible. It was considered more important to use the remaining time before closure for on-device assessment exclusively.

Another alternative to using behavioural measures of outcome was to use purely linguistic measures either on or off-device. For example, Mean Length of Utterance (MLU) was considered. On-device this was discounted as being too simplistic as symbols could represent more that one word or indeed utterance. Furthermore, users in this case were not using Easy Speaker as their primary means of communications and were therefore likely to be unrepresentative. Off-device the volume of data produced would be massive and highly specific even if sampling was used. What was need is a more holistic set of measures, namely the behavioural ones that were employed.

In a similar way grammatical analysis of utterances both on and off-device was discounted as this would be likely to be inaccurate as noted earlier most users in this population would be likely to produce ungrammatical utterances anyway due to economy. Improvements in grammatical complexity would likely to be measured over years rather than months. It is worth pointing out that on-device all users could construct multi-symbol utterances with ease; with the length and pure grammatical correctness dependant upon the available symbols and the ability of the user themselves.

The goal of this study was to support the notion that there was a tangible benefit for more able users other than the acquisition of a skill as discussed in the previous chapter and the ability to use Easy Speaker as a primary communication channel if the need arose. The fact that more able users displayed secondary benefits of using Easy Speaker is encouraging and has positive implications of long term use of Easy Speaker as a primary communication channel for the more severely Learning Disabled non-vocal. This is especially true in terms of motivation, memory, attention and social interaction.

As all users involved in the study were borderline vocal, but representative of the abilities and handicaps displayed by the majority of both the non-vocal and vocal LD population, one can conclude that typical improvements discussed would be likely to be generalisable. More importantly less able users should be able to construct multi-symbol utterances which are semantically meaningful. The implication here is that Easy Speaker might be used as a primary means of communication within a linguistically impoverished LD population.

It is not suggested that Easy Speaker use by the non-vocal LD individual would lead to their spontaneous use of voice and natural language. However, it is suggested that prolonged Easy Speaker usage would be likely to improve their communicative ability both on and off-device. That is, both verbally and non-verbally. For the more vocal user, use may lead to improvements in their residual vocal and language usage as occurred within the user group under study.

Only longer term studies over a number of years, with both vocal and non-vocal LD users, would be likely to confirm these conclusions. However, conclusions drawn from successes over the six month interventional study discussed are very encouraging. Furthermore, they lend support for the numerous informal studies carried out by Rostron and Plant et al 1992 onwards. Consistent improvements were made by all fours users, which can in effect be taken as successful positive replications in a clinical setting due to the incorporated single-subject with parallel replication design (Barlow and Hersen 1976). Combined with results from chapter nine, in which users were found to easily be able to control, to learn to use Easy Speaker efficiently, and to navigate a large and relatively complex hierarchy building up meaningful multi-symbol utterances this is particularly encouraging. This is especially true for individuals previously considered on such low ability that they were precluded from making used of complex, high demand ACD's in order to communicate their needs and wants *vocally*. Not only then should Easy Speaker be able to be used as a primary means of communication but it might also produce secondary benefits in language and cognition.

SUMMARY

This chapter has focused on whether there are any measurable off-device benefits of Easy Speaker usage over a prolonged period by application of standard clinical adaptive behaviour questionnaire items. Each user was rated by *two raters*, both prior to intervention and at six months. Overall all users showed a very highly significant improvement over baseline in adaptive measures of language and cognition as measured behaviourally by the ESPM questionnaire. In more detailed terms there were specific improvements in :-

❶ Expressive language
❷ Receptive language
❸ Social language development
❹ Initiative
❺ Attention
❻ Memory

Each users ratings in the above areas have been plotted and analysed and discussed individually. Within the final section of this chapter an overview of the findings as a whole has been presented. These results are supportive of the ethos of low demand ACD design which potentially leads to easier device operation and can lead to, and promote off-device improvements in language and cognition.

The implications for less able non-vocal LD individuals are favourable given the improvements shown by their slightly more able peers. Remeditaion implications for all however are very encouraging.

CHAPTER 11: Conclusions and future trends

The role of this thesis has been to follow the development of a new category of Augmentative Communication Device designed specifically for the linguistically impoverished of the Learning Disabled population. Development has been reported from theoretical, practical and evaluative perspectives.

The practical result has been to produce a piece of ACD software (Easy Speaker) which has been designed to take account of the abilities and handicaps of the intended user population. In developing Easy Speaker a hypermedia metaphor has been implemented and combined with high quality photorealistic symbol set linked to naturalistic digitised speech output.

A short-term skill acquisition study has found that real users can acquire the necessary skills to make productive use of the system as a communication device. A longer-term field study has found significant off-device gains in language and cognition in those users. Despite the main thrust of development being geared to end users, the opinions of maintainers toward Easy Speaker was found to be significantly better than expected.

However, as with any piece of research, strengths and weaknesses of the approach were revealed during investigation. In this case although overall results were encouraging, conclusions as to the implications for the current research, and potential users must be drawn. Independent research using Easy Speaker with a variety of user groups has given strong external validity to the overall ethos and design of the software itself. In the light of these conclusions and with newly emerging technologies future research directions and trends are also considered.

Introduction

Unfortunately an almost 'folklore' attitude toward modern technology and the Learning Disabled population still exists. With the view that cutting-edge technology and learning disability is a mix best avoided. This is reflected in many ACD's which continue to be made use of by only the most able due to their inherent complexity. However, as many researchers have pointed out, it is the most able who are rated as being adept in terms of communicative effectiveness. In short, they tend to get their message across, despite common linguistic and other handicaps, with or without the assistance of an ACD. This leaves a significant number of less able LD individuals who are precluded from benefiting from the use of many commercial ACD's either as a primary or secondary means of communication due to device complexity. Many earlier devices have been designed by electronic engineers, computer scientists or enthusiastic third parties to fulfil a specific need. Unfortunately many only meet the needs of the more able, or are never field trialed with real users. Newell and Alm (1994) sum the situation up well with the statement that :-

> "[often] designers have assumed that their users will have the characteristics of a 25 year-old male who has a PhD in computer science and is obsessed with technological gadgets rather than getting on with the job. Systems are a triumph of functionality over usability."
>
> Newell and Alm (1994)

Only very recently have systems which are simpler to operate, and therefore open to potential use by even the least able appeared commercially. As one might expect these are typically commercially driven products that have been produced as a result of a real need and because "they work". However, because "they work", commercial devices tend not to have been thoroughly, and objectively field trialed. Nor do they tend to be based around psychological, cognitive, or communication theory. In addition, the maintenance of such device by caregivers, or health care professionals, is seldom mentioned, and *never* researched.

Easy Speaker however has been designed and developed with the needs and abilities of *all* potential users in mind. Even the least able should be able to use the system as either a primary or secondary communication channel, or simply as a learning tool or recreational environment. Throughout development emphasis has centred around good, as opposed to compromised design. Through good design, synergistic interface metaphor, and use of leading-edge technologies the cognitive overhead imposed during device operation has been reduced to acceptable levels.

Comprehensive, well controlled, field studies have taken place to investigate whether real users can make productive use of the system. These have shown that over a relatively short period users can effectively learn to operate the system, and make significant gains whilst doing so. Longer term field studies have also shown significant secondary gains in residual language and cognition off-device when measured in terms of the users adaptive behaviour. This was suggested to be a result of the interface metaphor used combined with the use of high quality digitised speech.

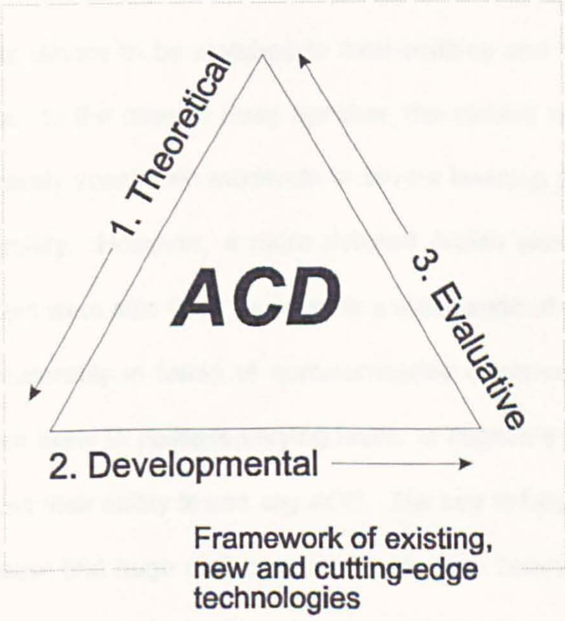
Maintainers, or caregivers, who might help construct hypermedia hierarchies for users rated Easy Speaker as being significantly better than expected. That is, most found it easy to operate and tailor for real users with a range of handicaps. Such acceptance is crucial for devices in the real world to be made full use of. This is especially true of service professions such as clinical psychology and speech and communication therapy services within the NHS. Typically such services are slow to uptake advances in new technology especially if over complex and time consuming to implement (Mansell 1992).

Given such favourable results these might be viewed as supporting the psychological, cognitive and communication theories on which the design of Easy Speaker is based. However, before accepting such conclusions one must collate, and review, the evidence from which such assertions are made.

CONCLUSIONS: A MODEL FOR SUCCESSFUL ACD IMPLEMENTATION

When collating, and reviewing, the evidence for the successful development, implementation, and monitoring of an ACD such as Easy Speaker it helps to provide a model through which the evidence can be focused, and conclusions drawn. Such a model might take the form of the three stage model for successful ACD implementation proposed by the author :-

Fig 11.1, A three stage model for successful ACD implementation



According to the model successful development should follow a three stage interactive process. The first stage should be to establish a theoretical base for the development of a new type of ACD. Second a developmental and refinement stage should develop the device and iron out any problems in the design and implementation. Finally, the third should evaluate, or experimentally test, the new ACD with real users. This final stage should produce objective data regarding on-device performance and off-device gains in areas such as language and cognition.

All three stages should be completed under a framework of existing, new, and cutting-edge technologies. For example, as available new technologies become available improvements offered should incorporated into new ACD designs. Existing ACD's and proven technologies should also be taken account of. The cyclical nature of the model is indicative of how the fruits of each stage should feed into the next infinitum ...

The development and implementation of Easy Speaker has followed this three stage model and will continue to do so. Having outlined the principles behind the model it makes sense to review Easy Speaker progress according to each of the three stages before attempting to draw any conclusions.

The theoretical stage

The starting point for any research involving real people is to know the population with which one is dealing. The development of a successful ACD relies on the correct identification of the potential user population for the device to be matched to their abilities and handicaps as Newall and Alm 1994 rightly point out. In the case of Easy Speaker, the system was designed for potential users that were non-, or barely vocal, with moderate to severe learning disabilities, and a variable level residual language ability. However, a more detailed profile was needed. As outlined in chapter three potential users were also likely to possess a wide range of communication disorders, and as a result vary considerably in terms of communicative effectiveness (Blackwell, Hurlburt and Bell 1989). They were likely to possess varying levels of cognitive and physical impairments that would be likely to affect their ability to use any ACD. The key in targeting such a population is the overriding need to realise that huge variations can exist, both between, and within individuals themselves. For example, there can be wide variations in attention and memory spans. Unfortunately such variation also precludes the use of standard control groups and application of the standard experimental model. One must also allow for the common occurrence of uncorrected visual and auditory impairments. As a result there are many variables that can affect both the theoretical and practical approach one takes to developing a successful ACD such as Easy Speaker.

The first concern was the need to provide low demand, easy access, to a potentially huge vocabulary of speech. Due to its obvious similarity with models of human knowledge representation and language processing (e.g. Rumelhart and McClelland 1986), hypermedia was suggested to provide an ideal low demand interface. In chapter four this notion was explored in some depth and its relationship to the practicalities of Easy Speaker development stressed. It was also tentatively suggested that by use of a synergistic hypermedia interface secondary gains in cognition may also result. It was suggested that if real users could learn to use Easy Speaker productively then hypermedia provided a successful interface metaphor. This, proved to be the case with some users achieving usability scores equivalent to unimpaired users as discussed in chapter nine. Many users also showed off-device gains in language and cognition. Such assertions, although subjective, are encouraging and were discussed in chapter ten.

The second, was to provide a symbol set to which users could relate. That is, many previous systems made use of abstract, or idiosyncratic, symbol sets that needed to be learnt by the user. Obviously for individuals with lowered cognitive abilities the ability to learn, manipulate, and make use of a highly abstract symbol set is questionable. Taking account of potential users' abilities within chapters three and five, the merits of abstract and more concrete symbol sets were explored. Although more concrete symbol sets do exist it was suggested that these can still require a great deal of effort on the users part to learn, manipulate and use effectively. Both previous research and common-sense semiotics suggested that photorealistic images of actual events, objects and people would provide the least demand, high productivity, symbol set. With some researchers arguing that many of the incidental pairings of product logos and the product itself should be seized upon and made use of within a symbol set (Reichle, Sigafoos and Remmington 1991). This idea was taken literary in Easy Speaker which allows photorealistic scanned images of around passport photo size to be used as symbols. For example, a passport sized picture of a real can of Coke could be used to indicate the user wanted a drink of Coke. The speech linked to the symbol could then be spoken so that a caregiver could fulfil the request, e.g. "Can I have a drink of Coke please ?". The conclusion reached was why should a user be taught an abstract symbol, when they already understand a semiotically correct symbol, namely the product logo. Again the fact that users could make use of Easy Speaker was taken as support for this notion. The idea of including the user, and items relevant to the user, within symbols further strengthened this notion.

Thirdly, a theoretical model of how users might interact with the system was needed. The model applied to design was the SSOA (Syntactic-Semantic Object-Action) interface model in relation to direct manipulation (Shneiderman1992). By application of the model this enabled the author to differentiate between the knowledge required operate Easy Speaker, and the knowledge user needed to know about the computer per se. For example, what happens when symbols are selected, and what happens when a key on the keyboard is pressed to initiate movement. As the overriding ethos was to develop a low demand, high productivity ACD this enabled the amount of information the user needed to know to be formalised and then reduced to a minimum during development. That is, the less complex the interface, the less knowledge the user needed of it, and therefore the less demanding it is to make productive use of. In addition it is also easier to learn, and potentially open to use by even the least able. Indeed by making use of an hypermedia interface metaphor with photorealistic symbols this helped reduce the knowledge required by the user to an acceptable level. Although not a true task analysis users needed knowledge of only around ten system related items to make productive use of Easy Speaker, e.g. that red boxes around symbols signified a link to another part of the hierarchy which contained more relevant symbols. The theoretical notion of low demand access through minimum task and system knowledge was supported in chapter nine in which users competence on set tasks was measured. Over a period of three sessions users measured efficiency and constructional latency data were plotted to reveal that they very quickly learnt to operate the system productively despite their handicaps. Their performance curves bottomed out rapidly suggesting optimum levels of learning and performance were reached. Such rapid gains in skill are suggestive of the system complexity being pitched correctly for the user group in question, with usability being high. In addition to being able to operate Easy Speaker analysis of tracking data suggested that all users could produce multi-symbol utterances in a truly generative fashion.

Finally, the contentious issue of whether to make use of synthesised or digitised speech output was addressed in chapter seven. Although from a software development point of view, synthesised, or text-to-speech, is simpler to implement, newly available technologies allowed for the use of digitised speech. Previously the main drawbacks of digitised speech were high machine demands and it's non-generative nature, or specifically concatenated words. Progress has been made in both these areas with more powerful readily available hardware, tighter and faster compilers, and the ability to store individual digitised words which could be linked on the fly to produce phrases. Fortunately Learning Disabled users are likely to possess a smaller working lexis which in real terms means that the number of symbols needed is smaller than those for an unimpaired individual for whom a truly generative system such as text-to-speech is possibly better. Many have accepted that synthesised speech involved different resources when listening to it. However this was ascribed to a shift in emphasis, as opposed to a shift, or increase in capacity, and that training, or 'tuning-in to the speech', could help compensate. By applying well researched models of natural language processing, such as the Cohort model (Marslen-Wilson and Tyler 1980), it was suggested that synthesised speech is more cognitively demanding. By applying this model it was suggested that for those with decreased cognitive abilities a form of output should be used that was least cognitively demanding on decoding. This was naturalistic digitised speech. Based on such theories it was decided that spoken output should be in the form of digitised speech for five key reasons :-

- It imposes a lower cognitive demand on decoding, or listening
- Models suggest that speakers, or in this case ACD users, reprocess their own spoken output
- Users are likely to interact with others with decreased cognitive ability
- High quality output may improve the users own residual speech as the ACD may act as a constant speech therapist
- Digitisation allows not only speech to be recorded, but also music, and other sounds. For example, a symbol of the users dog can be made to bark by recording the actual dogs bark

Support for the cognitive overhead imposed by synthesised speech was experimentally evaluated by asking unimpaired subjects to complete increasingly complex tasks which they were instructed to carry out in either synthesised or digitised speech. The result was that task completion times were significantly higher when obeying synthesised speech instruction even after practice 'tuning'.

similar to the Cohort model resulting in synthesised speech incurring a larger processing overhead. It is proposed that LD individuals might also employ this model, but that any increases cognitive load would be likely to hit them hardest due to already limited resources. Field support for the implementation of digitised speech comes from the ratings of caregivers after Easy Speaker intervention. In which they rated the majority of the users residual language as having improved. That is, both in terms of receptive and expressive language usage. In addition users themselves 'loved it' as it provided for not only digitised speech but also for favourite music, for pet dogs to bark and for a multitude of other sounds to be incorporated.

The developmental stage

Easy Speaker has attempted to put theory into practice over a number of years. Initial versions were completed around 1990, with various revisions the version used in this thesis was completed in 1994. However, as suggested by the three stage model, developments and refinements are still in progress with a new version currently being finalised.

Development has, in the main, focused on the abilities and handicaps of potential users and has been tailored to take account of the relevant theoretical premises discussed above. However, as outlined in chapter five, the user interface has remained more-or-less constant for around four years. Improvements as far as users have been concerned have been due to advancements in PC technology and development software. The most important change came when Easy Speaker for Microsoft DOS was re-written for MS Windows.

As the new system ran under MS Windows it was device independent. For example, it no longer had to run with a specific sound device, or graphics card. It could now use any hardware that worked under MS Windows. This meant that more people could use Easy Speaker on standard instead of non-standard hardware. It also meant that graphics could be in any quality up to true colour, or 16.7 million individual colours. In short symbols became photorealistic. Sound could now be up to CD quality stereo, although for reasons of disk space this was reduced to the minimum for acceptable quality. With readily available hardware and optimised C/C++ compilers from Borland lag was almost imperceivable when selecting hypermedia links, or performing other tasks.

Photorealistic symbols almost 'threw themselves' onto the screen. Furthermore, the system had been rewritten so that speed was unaffected by the size of the vocabulary, or hierarchy. Typical storage requirements were around 30-50 MegaBytes per user. In addition any pointing device which worked under MS Windows could now be used to control the cursor. This meant that touch screens, joysticks, trackballs, pen input and other devices could be used as a straight plug in re-placement for the mouse.

Overall from a users point of view the Easy Speaker interface was complete. From a technical and theoretical point of view it made use of a true hypermedia metaphor, used a photorealistic symbol set, and had high quality digitised speech output.

Although the main thrust of development concerned usability for users, maintainer acceptance was considered crucial. In this respect the system was designed to be simple to maintain as possible. Just as design guidelines existed for users, so too did they for maintainers. The simple scripting language that Easy Speaker operates on has remained constant since its inception in 1990, with the only changes being in the tools used to create scripts, symbols, and digitised speech. Again the shift to a MS Windows environment was crucial. This enabled standard system applications, or applets, supplied with the operating system to be called upon by Easy Speaker. As Easy Speaker now ran under Windows one could guarantee that these applets would be present. In addition to saving a large amount of development time, this enabled consistency for maintainers, and ensured that they already had high quality manuals, and access to support for each of those applets. It was assumed that they were also likely to be familiar with them thus decreasing the amount they needed to learn or re-learn in order to create Easy Speaker hierarchies for users. MS Notepad was used for editing scripts, Paintbrush for editing symbols, Sound Recorder for digitising speech, and File Manager for file maintenance. A more in-depth discussion of the maintainer features was discussed in chapter five.

The move to Windows also saw the construction of an on-line guide using MS Help which was constantly available to maintainers. This was a cut-down version of the paper based manual which included screen shots and other graphical aids. In short, it provided information on how

Easy Speaker worked, how to edit scripts, use symbols, and record sounds. The guide itself could be searched, sections copied to the clipboard or printed out for reference.

From the second version onwards a tracking system known as the Automated Response Tracking System (ARTS) was incorporated into Easy Speaker. This was seen as crucial as it covertly monitored every user interaction with the system, noted timings, calculated performance means and estimated future performance on-the-fly. This enabled the author and maintainers alike to concentrate on the user, rather than having to worry about recording details of their performance manually. ARTS also allowed for the user to make use of Easy Speaker in isolation without being overshadowed by a *helper*. The aim of providing objective data on user performance was fulfilled. The collation of objective data allowed areas of weaknesses to be concentrated on and improved, and areas of strength to be consolidated. Thus the hierarchy the user was working with could be modified to suit and any improvements noted. The complexity and accuracy of ARTS is considerable and was fully discussed in chapter five.

The views of actual maintainers, in this case university students involved with work with real users, toward the system are crucial for its success. These were measured by using the University of Maryland Questionnaire for User Interface Satisfaction 5.0L (QUIS). Responses were evaluated against the expected average. Overall response to the design and operation of Easy Speaker from a maintainer perspective was significantly better than expected. Results and conclusions drawn from the QUIS were discussed in-depth in chapter six.

The evaluative stage

The evaluative stage concentrated on the ability of real users to operate and make productive use of Easy Speaker, and to measure any secondary gains in language and cognition. In order to carry out evaluations in an objective a way as possible bespoke test materials and questionnaires were constructed.

The first of these bespoke test materials was the Easy Speaker Assessment Tasks (ESAT). ESAT consisted of a set of eight utterances which users were required to construct and output using Easy Speaker. These increased in complexity and required users to select symbols, navigate hypermedia links, and output completed utterances. Whilst completing each task the tracking system, ARTS, recorded all interaction with Easy Speaker and logged timings for each event. Users completed the ESAT tasks once each week over a three week period. The use of standard tasks and highly accurate performance measures allowed for progress to be monitored objectively. Key measures were constructional latency and cursor efficiency. An estimated Words Per Minute (WPM) score was also calculated. The development and application of ESAT tasks was more fully discussed in chapter eight.

Data from actual users suggested that all could learn to operate Easy Speaker with relative ease, over a short period of time. All showed characteristic learning and performance curves with diminishing improvements in performance as time spent on the system increased. This suggested they were reaching their natural optimum performance levels relatively quickly. From this one can infer that theoretical notions of developing a low cognitive load system were substantiated. That is, specifically by application of the SSOA model, use of a hypermedia metaphor and photorealistic symbol set. All users by their third ESAT session had surpassed the sample norm. In addition to making significant performance gains over relatively few sessions, some users even approached, or in one case, surpassed the norms generated by an experienced unimpaired user.

Data on efficiency measures, i.e. cursor movement, also rapidly improved. Again by the third session all users had surpassed the sample norm, and some were approaching the efficiency of unimpaired users. This suggests, that users were honing their skills, making use of pre-planned strategies, and drawing upon previous experience. Again all users made significant gains in efficiency over ESAT sessions, until they reached optimum efficiency around the third session. Performance improvements are likely to continue but according to the law of diminishing returns.

As one might expect, based on improvements in constructional speed and efficiency, the estimated number of words per minute also increased. This measure was only included for completeness, as it must be remembered that Easy Speaker symbols can represent an unlimited number of digitised words in the form of a complete monologue. Obviously, as a result in some instances WPM estimates may be elevated, or for that matter reduced.

A combination of decreased constructional latency and increased efficiency in all users suggest that they could successfully make use of Easy Speaker despite their cognitive, linguistic and other impairments. Although the particular user sample were moderately learning disabled there is little reason to suppose that one cannot extrapolate these findings to those with a more severe handicap. This conclusion is based both upon the theoretical design of the system and upon the ease with which all users got to grips with it. Independent research by Rostron, Plant and Hermann (1994) in which a severely Learning Disabled individual made productive use of Easy Speaker supports this notion. Crucially no users were given formal instruction as to how to use Easy Speaker, which control methods to use and what navigational strategies might be most effective. For a full presentation of results and discussion see chapter nine.

The second of the bespoke test materials was the Easy Speaker Progress Monitoring (ESPM) questionnaire. This questionnaire was primarily concerned with measuring adaptive behaviours in relation to language and cognition. Items on the questionnaire were taken from the AMMR Adaptive Behaviour Scales (ABS) and NFER Functional Performance Record (FPR). These were grouped into six domains :-

❶ Expressive language
❷ Receptive language
❸ Social language development
❹ Initiative
❺ Attention
❻ Memory

It was considered *better* to assess off-device improvements in language and cognition using adaptive measures rather than attempt to use formal test measures. Improvements in adaptive behaviours and language are more relevant to the individual than a slight increase in any test performance measure.

The ESPM questionnaire was given to two of each users immediate caregivers prior to the start of Easy Speaker intervention in order to establish a baseline. After six months of Easy Speaker use the same questionnaire was re-administered. On the second occasion however the first responses for that caregiver were reiterated, and they were then asked whether they thought the user had become better or worse in relation to their previous response for that item. This allowed for a very sensitive measure of improvement. It also allowed for inter-rater agreement to be evaluated. The application of ESPM questionnaire was more fully discussed in chapter eight.

Data from actual users suggested that no user worsened in any of the six domains. On average each user significantly improved over baseline in three areas, as rated by both raters. Improvements were most frequently noted in expressive language, initiative, attention and memory. All improvements however were *significant* to the individuals themselves, and given past histories to their caregivers. Inter-rater agreement was around average as one might expect given the subjectivity of the questionnaire items themselves. For a full presentation of results and discussion see chapter ten.

In terms of evaluation Easy Speaker proved successful. Users could learn to make productive and efficient use of the system rapidly suggesting the correctness of the holistic approach. From this one can be fairly confident that more handicapped users would be able to make use of Easy Speaker as either their primary or secondary communication channel. Predicted off-device improvements were also demonstrated, with users improving in many key adaptive areas related to language and cognition. This would suggest that Easy Speaker also can have secondary benefits as a language and cognitive facilitation tool.

Framework of existing, new, and cutting-edge technologies

During all three stages of Easy Speaker development attention was paid to currently available and emerging technologies. Initially this took the form of a review of available ACD solutions, which continued up to the completion of this thesis. With device reviews continually being added as new technologies emerged. A representative range of devices were reviewed in chapter three, in which their relative strengths and weaknesses were stressed.

FUTURE TRENDS

Software advances

Despite Easy Speaker being a success this was not an unqualified one. Although real users could use the system productively and showed many of the predicted secondary gains in language and cognition, maintainers found the system to be quite demanding to maintain initially. As they became more experienced they found the system progressively easier to maintain. When considered expert they rated the system as being better than average overall, as assessed by the Questionnaire for User Interface Satisfaction (QUIS) in chapter six. However many still expressed concerns over the need to learn the scripting language and the high demands on organisation of files. For example, having to check on filenames for symbols and digitised speech clips, which were then inserted into scripts verbatim proved tedious. When constructing large hierarchies they found that spelling errors in scripts could be hard to pin down and the overall organisation could be lost if concentration slipped. That is, they were paying more attention to constructing scripts as opposed to concentrating on organisation per se, with correct structuring of material being beneficial to the user.

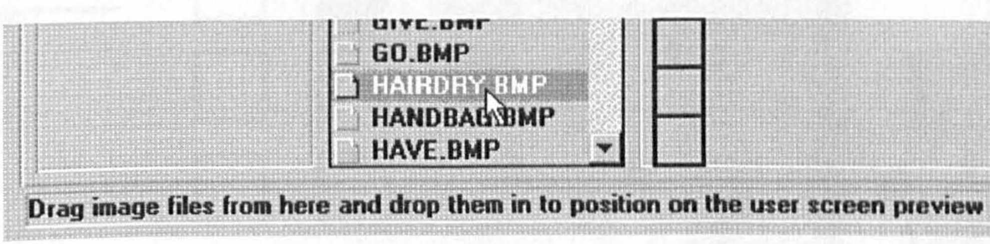
This is a potentially worrying issue when one considers that the maintainers involved within the QUIS study had the support of the author when required. Obviously in the field any ACD maintainer will be isolated with a manual, and on-line help system. If they're very lucky they may have telephone access with a support provider. Such concerns have led to a new version of Easy Speaker being produced which keeps the same successful user interface and tracking system (ARTS) but has a new front end for maintainers. The underlying scripting language remains but maintainers never see it, or even for that matter, know of it's existence !

The whole scripting language has been given a fully object-oriented graphical front end which lets maintainers simply drag symbols from a file manager and drop them into position on the screen where they want the user to see them. To associate a digitised speech file with the symbol they simply click on it to record one, or choose an existing one. All symbols and digitised speech can be previewed before being selected. Symbols can easily be reorganised or deleted by dropping them on a Macintosh style trash can. Thus the development of the hierarchy becomes simple, with the scripting language *hidden* from view, and the handling of a large number of symbols and sound files simplified.

In common with previous versions, Easy Speaker Pro for Windows, continues to make use of common system applets. These include MS Paintbrush, Sound recorder, and Help etc... New features include the ability to call, and make use of third party software tools. For example, Easy Speaker Pro can call graphics packages such as Paintshop Pro when a maintainer wishes to scan a picture in, or alter the colour balance. More complex third party software can also be called upon when required, e.g. a video grabber for capturing live video feeds.

In addition to the enhancement of on-line help the screen designer makes use of status bar tips. That is, when a maintainer moves the mouse over a given interface object they are informed as to its purpose by a textual description on the status bar :-

Fig 11.2, An example of the new status bar tips



The new object orientated maintainer interface is shown overleaf along with a description some of its features.

Fig 11.3, The new object orientated maintainer interface of the latest version of Easy Speaker

A standard menu bar allows files to be loaded, saved, and editied individually. In addition system applets can be called, third party tools started, and on-line help referenced

The file list box lists the files in a currently selected directory. As each file is clicked on it is previewed. If the maintainer wants to include the selected symbol, file or link on the current screen they simply drag it over to the postion they want it to appear in with the mouse and then drop it

When the preview is double clicked on the link shown is activated so that the asscoaited screen is displayed, or if a speech symbol the assocaited media is played

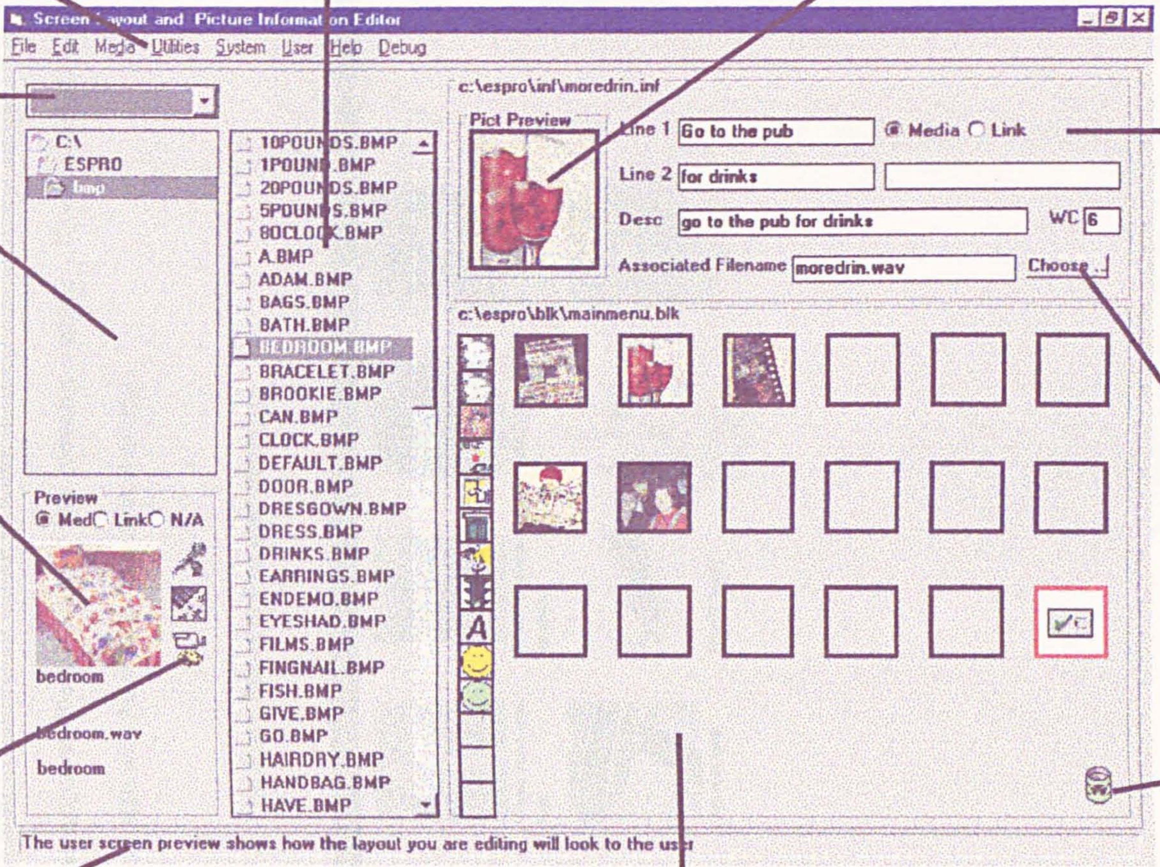
When a symbol, or link, from the user preview is double clikced on any information fo it is displayed her along with larger preview of the symbol. Information such as the text t appear under the symbol car be entered here, as can details of the digitised speecl content and length in words. If the symbol is to act as a link, the link radio button can be checked. Similiarly if it is to be a speech symbol then the icon button can be checked. To choose the destination of the link, or piece of digitised speech, MIDI score, video clip, or animantion the "choose ..." button is clicked on

The drive list box allows maintainers to select symbols, speech etc.. from different disk drives

The diectory list box allows maintainers to select the current directory on the drive choosen.

As each file in the file list box is clicked on they are previewed here if they are a symbol. Any exisiting information as to the type of symbol, i.e. link or speech, along with the text appearing under the symbol is also shown

If the maintainer wants to preview a peice of digitised speech they can drop the file onto the microphone icon. For a MIDI music score onto the control desk, and for a video clip or animation on to the video camera



When the choose button is clicked on the maintainer is presented with a standard fil manager from which they cai choose the destination scree for the symbol if it is a link, oi the type of media assocaited with that symbol, e.g. a piece of digitised speech

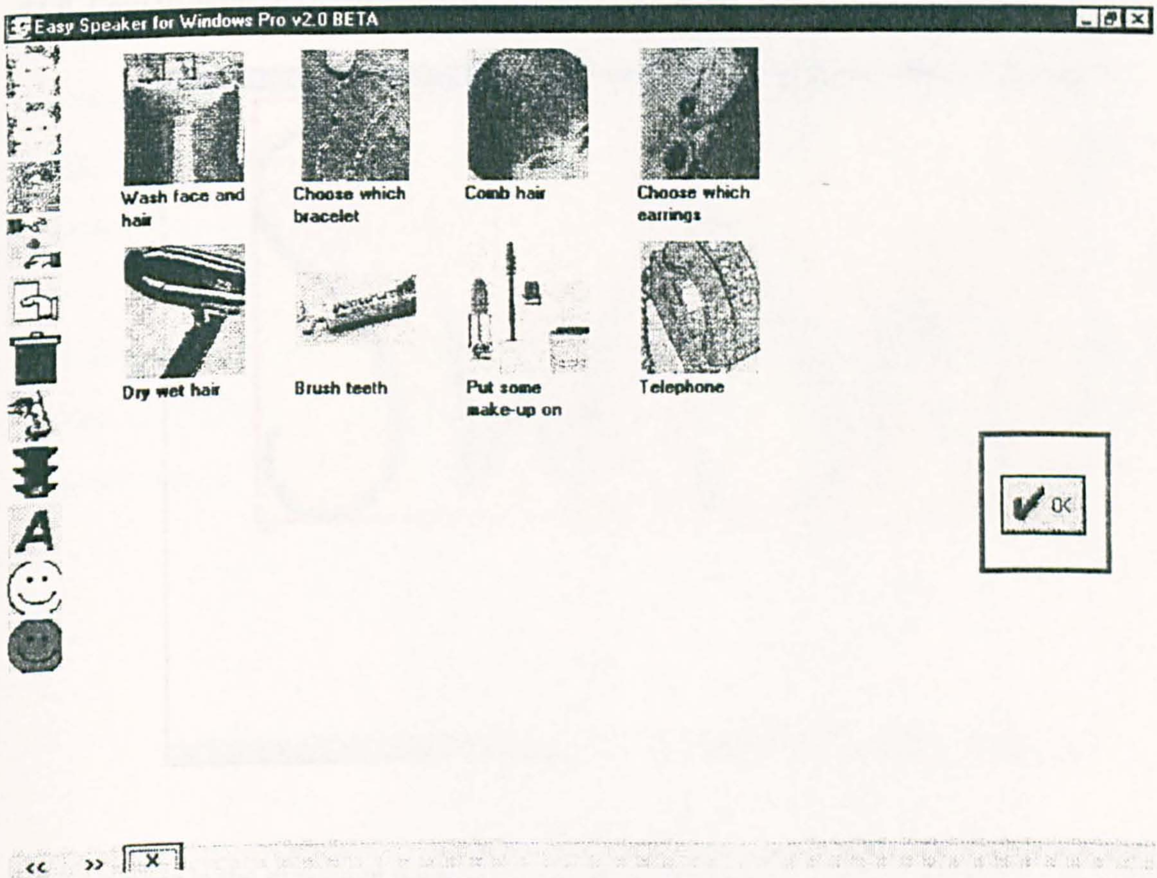
Symbols which are unwanted can be dragged from the usei preview and dropped in the bin.

A status bar changes to provide a textual description for the interface object the maintainer caurrently has the mouse cursor resting, or travelling, over

A preview of the user screen is shown here. Symbols are simply dropped into position from the file list box. Alternatively they can be dragged and dropped from any position in the user preview. Thus the layout can easily be rearranged. Links and speech symbols are indicated as normal. With hypermedi links being signified by a red bounding box. However text under speech symbols is not previewed.

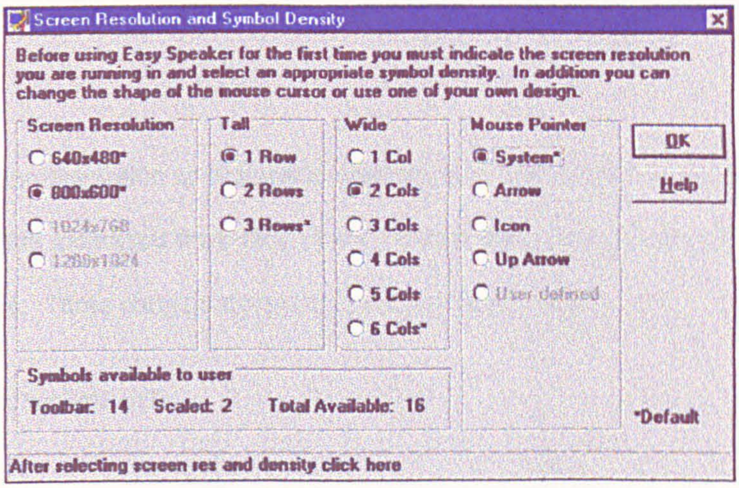
As far as the user is concerned Easy Speaker still operates in the same manner, bar that there are now 18 full size symbols instead of 8, and 14 smaller toolbar size symbols instead of 11 by default. Although users did not complain about symbol density it was a recurrent concern for maintainers who felt they would like to include more symbols on any one screen. Easy Speaker Pro provides around a 60% increase in symbol density over the previous version by default. However, this doesn't lead to decreased clarity as Easy Speaker Pro runs in a 800x600 pixel window where possible as opposed the 640x480 of the previous version. The size of each symbol is still roughly passport size with no visible reduction is clarity.

Fig 11.4, The Easy Speaker Pro user interface

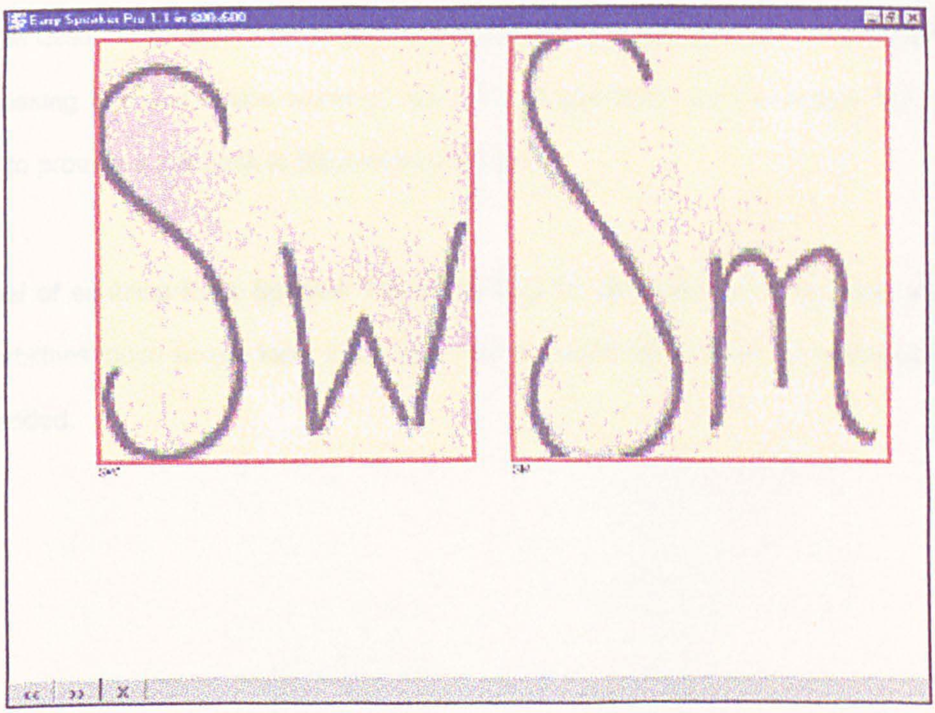


In addition to increasing the default number of symbols available to the user, this new version also allows for dynamically scaleable symbols. That is, symbols are scaled 'on-the-fly' before they are displayed on each user screen. Thus screen resolution and symbol density is accommodated.

Fig 11.5, Easy Speaker Pro allows for several screen resolutions and symbol densities



11.6, Easy Speaker Pro displaying just two large symbols per screen



The increase in the maximum number of symbols available at any resolution should allay the concerns of some maintainers who thought that the original symbol density was too low. Other maintainers have welcomed the change as they can now use variable size and density symbols which can help overcome uncorrected visual problems in some users. In addition for those of very low ability a smaller number of symbols is also helpful. Some communication therapists have found this a god send when using symbolic communication where one or two symbols are all a user requires when first learning the abstract nature of language, e.g. yes or no, please and thank you etc...

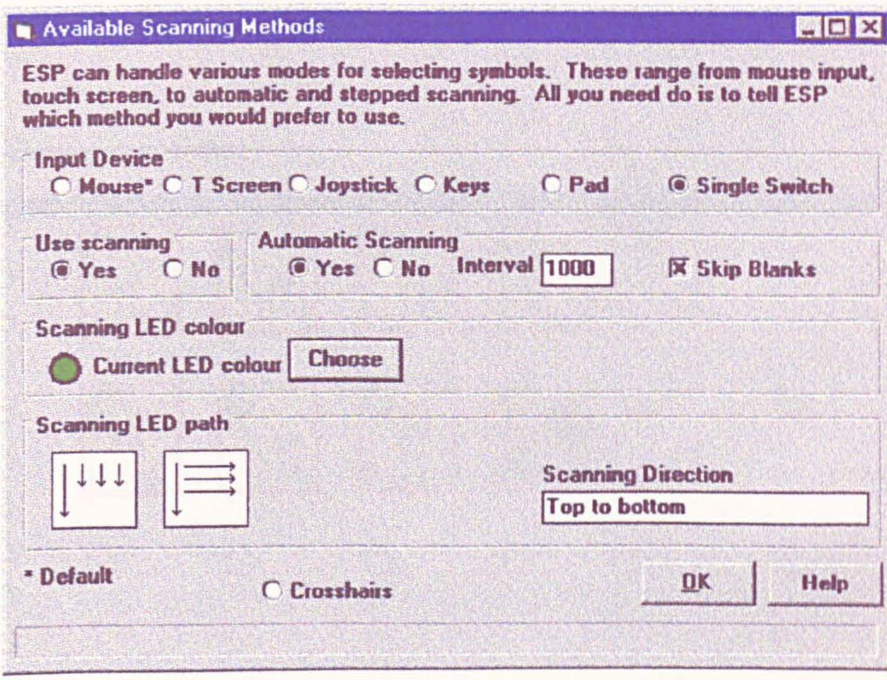
Regardless of screen resolution or scaleable symbols size the number of selected symbols that can be displayed has increased from four to six, so that the length of utterance on display can more representative. Three completely new buttons have been added.

Fig 11.7, New back, forward and undo buttons

The buttons, from left-to-right, provide a way for a user to jump back a link, go forward a link, and undo the selection of a speech symbol that has been added to the utterance under construction by mistake. Some users expressed the desire to be able to undo selections and step back through links already taken. Although increasing demand, these features are perhaps welcome, and decrease the onus on maintainers to provide a link back to the previous screen.

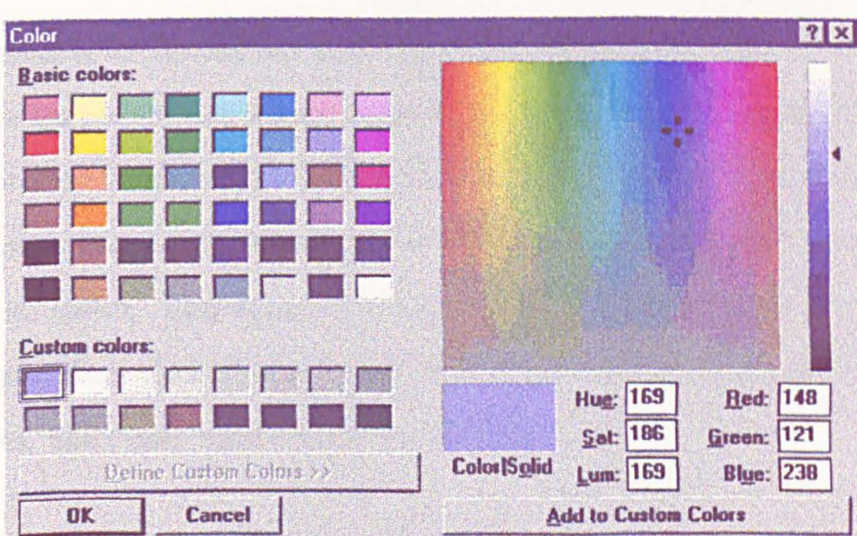
With the goal of enabling Easy Speaker Pro to be used by the least able and those with gross physical disabilities touch screen input and single switch control via a variety of scanning methods have been added.

Fig 11.8, Selection of input and scanning method



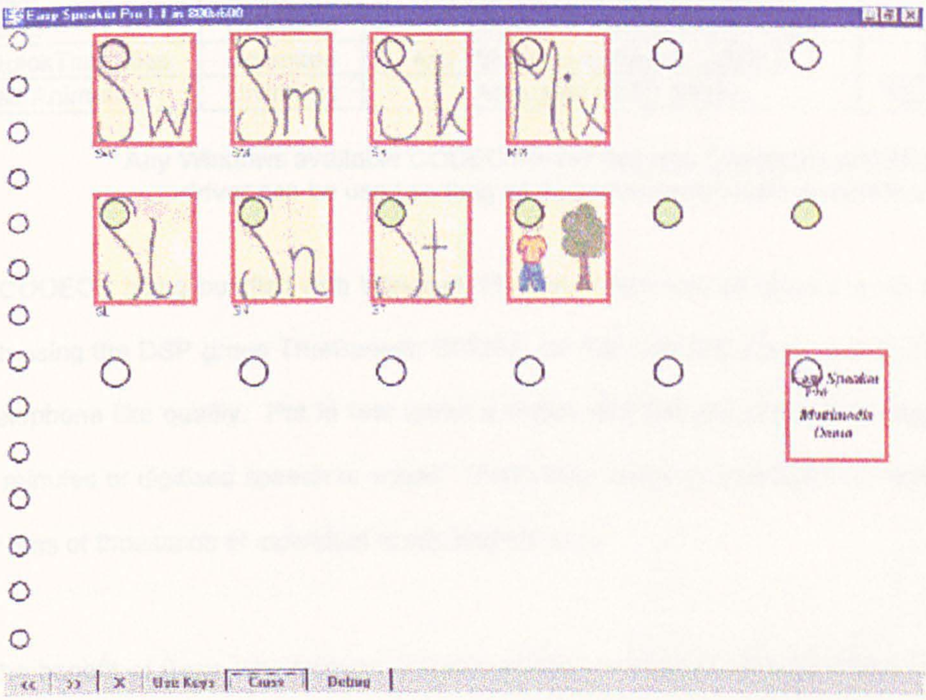
So advanced is the new scanning system that even the simulated LED lights colour can be selected so that the user can best see the scanning rows and columns of LED's.

Fig 11.9, Selecting the colour of simulated LEDs used in the new scanning interface for single switch control



The default scanning method (row and column) is based upon lighting each row of LEDs up in turn until the user hits a switch to stop them, the LEDs will then scan across that row one light at a time until the user hits the switch again and selects the symbol. That is, each row is illuminated and turned off in turn to give the impression of a moving line of LEDs, then each LED of a particular will scan across the screen. In addition to being able to change the colour of the LED the size and shape of them can also be changed if needed.

Fig 11.10, Scanning in operation



Since Easy Speaker for windows was written around 2 years ago, improvements in readily available hardware has been making frenetic progress. With Easy Speaker Pro now being 32 bit running primarily under Windows 95, 3.1 with Win32S, or OS/2 Warp. With this increasing power it has been sensible to utilise the technology to provide more than photorealistic symbols and high quality digitised speech. Easy Speaker Pro can now link various other types of media with any symbol. These include video clips, animation, MIDI music files, and Apple QuickTime video and animation files. Specifically, supported formats include :-

Fig 11.11, Media formats supported by Easy Speaker Pro

Media Type	Supported Length	Sub-formats	File Extension
Digitised sound	Unlimited	Any Windows available CODEC *	WAV
Video clips	Unlimited	Any Windows available CODEC *	AVI
MIDI music	Unlimited	4 to 256 instruments/Wave Table	MID
Apple QuickTime files	Unlimited	Any Windows available CODEC *	MOV
AutoDesk Animation	Unlimited	Animatior or 3D Studio	FLI or FLC

* Any Windows available CODEC means that any COmpressor/DECompressor driver can be used so long as its on the system and available to windows

With new CODEC's being bundled with Windows 95, video clips can be played up to full screen and speech using the DSP group TrueSpeech CODEC can be recorded using up only 1k per second with telephone like quality. Put in real terms a floppy disk has the capacity to hold approximately 24 minutes of digitised speech or sound. Potentially, even an average size hard disk can hold many tens of thousands of individual words and phrases.

Although the benefit of these new formats is untested they do provide another realm of possibilities for users. Not only can they use Easy Speaker as a primary or secondary channel of communication but they might also use it as a recreational environment. The ability to show and discuss a video clip with a friend is intriguing, as is the ability to play real music. In commercial devices such support is still years away due to prohibitive development costs. Unlike in the past the major cost of an ACD is now represented by software as opposed to the specialist hardware previously required. Their has been a paradigm shift toward software based ACD's running on standard hardware. Logically it is far better to concentrate of developing high quality software for industry standard platforms than on specialist hardware, as hardware can almost be 'left to take care of itself'. That is, real world computing needs and market forces will drive down machine cost and increase machine power whilst decreasing physical size.

Other key areas of Easy Speaker have remained unchanged. The most important of these is the tracking system ARTS which still continues to track the actions of users very accurately. This track is then available as a log file which can be read in to any spreadsheet and analysed accordingly. The only change being that this is created in both a Comma Separated Value (CSV) file and a Microsoft ACCESS database (MDB) format file.

It is not solely advances in operating systems and the Easy Speaker software itself that has improved usability for users and maintainers alike, and expanded the range of possibilities. New hardware has meant that such features can easily be implemented in real-time for acceptable programming effort and monetary investment.

Hardware advances

With the majority of hardware based ACD's still costing many thousands of pounds they are increasingly looking like poor investments. This has largely been due to increasing power of smaller and ever cheaper generally available computer hardware. At the forefront of this drive is the Windows/Intel, or WINTEL (WINdows and INTEL), alliance to produce faster and ever more affordable IBM PC compatibles. Software based ACD's such as Easy Speaker and its commercial counterparts are set to dominate. With operating systems such as Windows offering device independence one can easily see why this should be so. Such independence allows developers to rightly assume that their software will run on any device that can run an operating system such as Windows. In Easy Speakers case all that is needed is a Windows based machine with a sound card. This means that it can run on IBM PC compatibles, DEC Alpha's, PowerMAC's, and UNIX boxes. In the latter cases by virtue of appropriate versions of Windows 95, NT, or emulators. Thus development costs are reduced, and the range of equipment that the ACD can be used with is vastly increased. If a given user requires a smaller physical package to be more portable, then a smaller but equally powerful machine can be purchased. In a residential setting they may use a desktop with the same software and available vocabulary. It is far cheaper, and more productive to purchase more software than it is to buy specialist hardware that cannot be used for other purposes. The PC offers the ability to be able to perform other tasks, rather than being a solid state ACD. For example, the user can use it for recreational games, writing, drawing etc... In addition to which the ACD software itself can be upgraded with better version, or dropped altogether it does not suit a given user. The possibilities are endless.

In terms of readily available hardware which can run Easy Speaker and some of its commercial counterparts Pentium notebooks with touch screens and audio capabilities are now available for around £2,500, with 486 variants for around £1,000. Touch screens can be added as a retro fit for around £300. As computers are now treated as commodities, with chips being bought and sold as such, prices drop literally weekly, with impending chip releases speeding this downward spiral in pricing. This can only be to the benefit of the ACD user.

To give an example, currently the best piece of hardware [in the authors view] to run a software based ACD such as Easy Speaker on is the Fujitsu Stylistic 500, or something very similar.

Fig 11.12, The Fujitsu "Stylistic 500" pen and touch screen based mobile computer running MS Windows

Weighing in at just 2.5 lb., with a 486 DX2 CPU and giving over 6 hours of battery life the Stylistic offers the potential to be an ideal base for a software ACD. At around £2,000 it has a touch sensitive screen and in-built sound card capable of both recording and playback of digitised sound/speech and generating MIDI music. It also has the power to play video clips and animation's with ease. If fitted with a 32 Mb PMCIA III solid-state storage card it can offer instant access to photorealistic images and digitised sound/speech. If running MS Windows 95 and using the DSP groups TrueSpeech CODEC one could store around 9 hours worth of average quality digitised sound/speech. To realise the enormity of this achievement this casual statement needs to be put in prospective. A 32 Mb of PCMCIA flash memory card could hold around :-

In terms of speech

- 9 hours, or 546 minutes, worth of digitised sound/speech recording
- 33,000 individual digitised words at a rate of 1/sec

In terms of symbols (without JPEG compression)

- 7,281 - 16 colour speech or hypermedia link symbols
- 3,810 - 256 colour speech or hypermedia link symbols
- 723 - 16.7 million colour, or photorealistic, speech or hypermedia link symbols

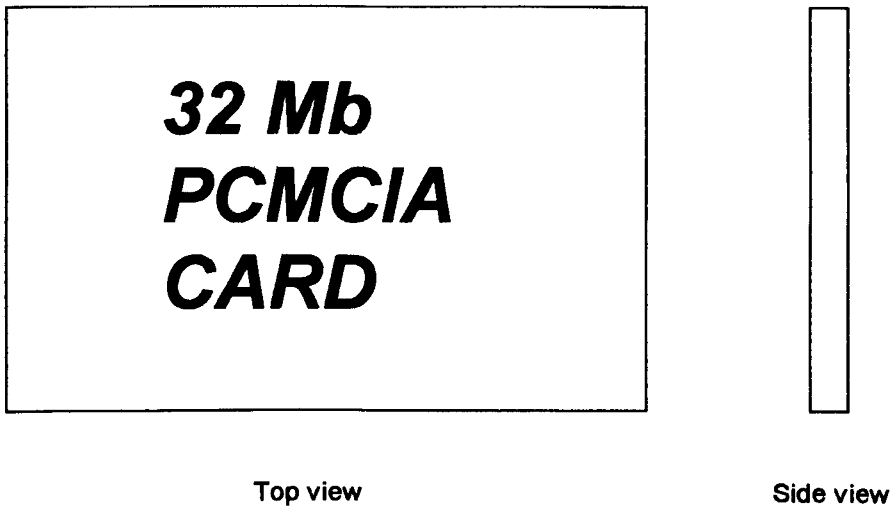
With JPEG image compression assuming an 80% compression ratio

- 3,615 - 16.7 million colour, or photorealistic, speech or hypermedia link symbols

This is even more remarkable considering the size of such cards.

Fig 11.13, A 32 Mb PCMCIA flash memory card drawn to scale

Dimensions: Length: 85 mm, Width: 54 mm, Depth: 5 mm, Weight: 35 g



In concrete terms one 32 Mb card could store enough digitised words to be equivalent to that of an average pocket dictionary. Certainly enough for normal conversation. These words or phrases could be represented by up to 7,281 high quality symbols and links. The potential then for such an ACD using currently available hardware is clear with card and machine capacities increasing almost daily.

ACD's are no longer constrained by under-powered hardware using low quality speech synthesis output, and abstract symbol set. Further developments in hardware, operating systems and ACD software can only improve the communicative effectiveness for those otherwise considered incapable of operating previous generations of ACD's.

Corroborative studies using Easy Speaker

Since the original MS DOS based Easy Speaker software was written there has been a growing interest in the approach, the philosophy behind it, and the uses to which it can be put (Rostron and Plant 1992). In testimony to the software itself and to the ethos behind it some of the more notable and relevant studies are outlined below.

Rostron, Plant and Hermann (1994) for example reported on a 22 year old man (known as R.) with severe Learning Disabilities and severe communications problems. R.'s communicative ability, despite attempts to teach him Makaton, was very limited. Unfortunately he had made no attempt to use Makaton in any way. His capacity to learn was described as "minimal" according to his caregivers and given background history. Given his background the prognosis for Easy Speaker use was apparently poor. With the help of one of the authors who had known R. for two and a half years a customised hierarchy was constructed which contained relevant pictures, speech and sound, e.g. pictures of a recent holiday and pieces of music were used. Given R.'s abilities the purpose of this material was to :-

- Provide a means of assessing the user's capabilities
- Devise an environment where communication skills could be developed
- Offer a novel entertainment and recreation environment

Specifically data was gathered on each of the following criteria :-

- Can the user recognise and identify pictures and symbols ?
- What aspects of language can be understood and responded to ?
- Are there any signs of productive language use ?
- To what extent is the users attention held by the particular environment and how long does it remain of interest ?
- Is the situation sufficiently motivating to warrant developing material for other specific learning tasks ?

R. had a recognised, but uncorrected visual problem, so it was informative to discover that he could recognise and identify pictures and symbols correctly, albeit at a close distance. For example, he easily recognised pictures of where he'd been on holiday when they were used as symbols on screen. Given that he could recognise the chosen symbols next the authors used Easy Speaker to assess what aspects of spoken language he could understand and respond to. Previously it was uncertain as to what R. could actually understand and what he chose to ignore. This consisted of asking R. to point to various symbols on screen which were then output as digitised speech or music. Over a number of sessions the complexity of the requests was increased and R. was shown how to navigate between screens of symbols using hypermedia style links. This culminated in R. being able to answer simple questions such as, "What did you have to drink in the café on holiday ?", "Who was there ?" etc... using Easy Speaker and the symbol set provided. He produced multi-symbol utterances such as "I drank lots of - tea" and "I went to the - café with T." demonstrating a degree of generative usage. In this case each symbol represented multi-word digitised speech utterances. R. also demonstrated the ability to navigate between screens following the structure of the hierarchy. He was particularly fond of Kylie Minogue at the time and so a picture of her was turned in to a symbol and buried within the hierarchy. When selected an piece of Kylie Minogue music would be played at which point R. would be come excited and highly motivated. Despite the symbol being buried three or more levels within the hierarchy R. consistently managed to activate it in order to listen to the music. This suggests that he could learn given an appropriate setting and level of motivation. This suggests that R. could have produced multi-symbol utterances from more than one screen despite the originally poor prognosis. In terms of motivation and attention R. remained highly motivated and attentive during each session which lasted for up to an hour. Indeed he had to be 'physically prised' from the computer on more than one occasion suggesting his diagnosed short attention span was almost irrelevant. Such positive results prompted the authors to construct more material for R. through which he could both learn and communicate. Unfortunately R's requirements dictated that he would need a highly portable device such as the Fujitsu Stylistc discussed earlier, and was hindered by the desktop PC used. In 1992-1993 when the study was conducted no such machines were readily available. However, it demonstrates that the user interface of Easy Speaker was suitable and that those with severe Learning Disabilities could use it as their primary means of communication.

Rostron and Kinsella (1995) made use of Easy Speaker not as a Augmentative Communication Device but as an aid in improving pronunciation in unimpaired Italian language learners. Rostron and Kinsella let half of the learners have access to Easy Speaker with an appropriate set of symbols and digitised Italian words and phrases. The other half carried on using traditional methods. The hierarchy used contained symbols such as a digitised photo of a "cup" with the English text "cup" underneath. When selected the Italian word for cup would be output as digitised well pronounced Italian speech. The goal was for learners to try to imitate this high quality output. When both groups were assessed blind for pronunciation quality by course lecturers the group who used Easy Speaker were found to be significantly better. Although apparently removed from using Easy Speaker as a VOCA or ACD this does testify that using high quality digitised speech is the correct approach regardless of language. That is, it has the inherent capacity to improve and shape the users own language, or in this case their Italian pronunciation skills. Obviously this is something that synthesised speech cannot offer. In language impaired individuals any improvement in residual language would be welcome. Such a finding supports the notion of improved expressive language off-device in Learning Disabled Easy Speaker users discussed in chapter ten.

Rostron and Gimson (1995) used Easy Speaker in a similar role in order to teach difficult phonemes symbolically to a severely hearing impaired 8 year old girl (known as M.) with marked pronunciation and intelligibility problems. Other than her deafness M. was otherwise unimpaired and of normal intelligence. Rostron and Gimson aware of the particular problems deaf children have in pronouncing particular words and discriminating between them constructed a hierarchy which contained particularly difficult words. For example :-

Fig 11.14, An example of difficult to pronounce phonemes and their symbols as used in Easy Speaker with a severely deaf child

Phoneme	Words
SM	SMell, SMack, Smile (from left to right)

Each segment of digitised speech that was generated placed stress or empathises on the phoneme in question. M. was encouraged to copy these and other words as they were output. All sessions were audio taped and compared with baseline so that improvements could be assessed. M.'s teacher noted a significant improvement in pronunciation after Easy Speaker use. An independent deaf teaching expert also rated an improvement based on the audio taped sessions.

In its intended role as an ACD a dysphasic stoke patient made use of Easy Speaker as a means of communication with a good degree of success in the home environment (Rostron, Plant and Ward

1996). RL. a 61 year old man used Easy Speaker for regular activity sessions during a four week period in his own home. He was encouraged to use Easy Speaker constantly during this period with the hierarchy constructed for him. Prior to Easy Speaker use he had no computer experience. In this study Rostron et al assessed :-

- Learning measured by
  - Speed of use - using total time taken to construct and reproduce an utterance and predicted words per minute score on set tasks
  - Accuracy of use - using total number of button presses and total number of mickeys the mouse moved in constructing utterances
  - Delayed recall - recalling the symbols selected during each session
- Overall familiarity
  - Number of symbols located - locating a number of randomly chosen symbols within the hierarchy
- Utility of the ACD
  - Independent use - independent use of Easy Speaker between formal sessions
- Changes in functional communication
  - A comparison of scores on the "Modified Communicative Effectiveness Index" pre and post-intervention

In terms of time taken to produce set utterances over nine sessions there was a highly significant improvement over the project period. Over the nine sessions there was a gradual increase in the predicted words per minute output, although this was not significant. In terms of accuracy there was a significant improvement in the number of button presses over the nine sessions, i.e. less button presses to achieve the same result. In terms of total mouse movement in mickeys there was a very highly significant improvement between sessions. In terms of delayed recall when questioned RL. recalled significantly more words output during a session as the number of sessions increased. That is when questioned after each session.

When overall familiarity was assessed by asking RL. to navigate to certain symbols during the final session he attempted 31 of the 40 tasks before he was judged too tired to continue. Of the 31 items 7 were completely correct, 12 were attempted but were not ultimately successful and 12 were not known and not attempted.

In terms of utility RL. used the computer on only four occasions in-between the formal sessions. It is possible that this was due to RL's general condition. Output from the occasions when it was used however showed that meaningful utterances had been generated.

As regards changes in functional communication a second post-intervention questionnaire was not completed due to lack of informal ACD use. Lack of informal ACD use was attributed to RL's premorbid state and general lack of motivation to attempt to communicate. In the words of his wife, "he was never a great talker". One other family member could almost 'read his mind' and on many occasions pre-empted ACD use by 'talking for RL.'.

Despite the mixed findings with this dysphasic user, the fact that he could control Easy Speaker, improved with practice, and learnt parts of the hierarchy is encouraging. One must bear in mind this was despite his weak condition and general stereotyped patterns of communication. Given a stronger physical state and better motivation he might well have used Easy Speaker in order to communicate exclusively.

Finally in papers by Plant and Rostron (1996) and Rostron and Plant (1996), I outlined the importance of the caregiver interface as well as that presented to the user. Stress in these articles was placed on the importance of how easy any ACD should be to maintain for a given user. The new features of the latest version of Easy Speaker Pro were discussed, e.g. drag-and-drop setting up of user screens. In time pressured environments this was seen as crucial and was considered as being of equal importance to good design of the user interface. Ease of caregiver use was stressed as many commercially available ACD's pay scant attention to this and are time consuming and difficult to set-up and maintain. Overall feedback to the papers was favourable and numerous copies of Easy Speaker have been sent out and downloaded from our Internet site as a result.

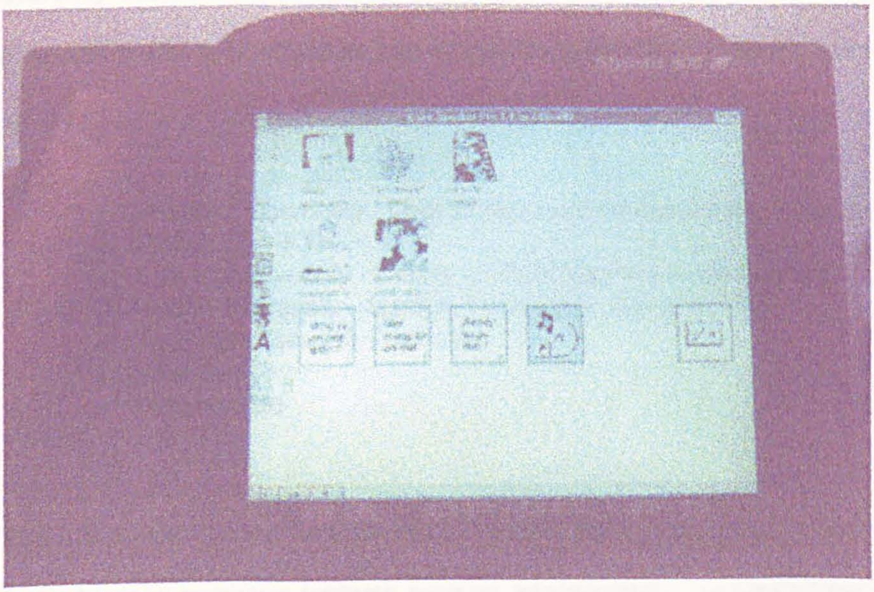
Future research

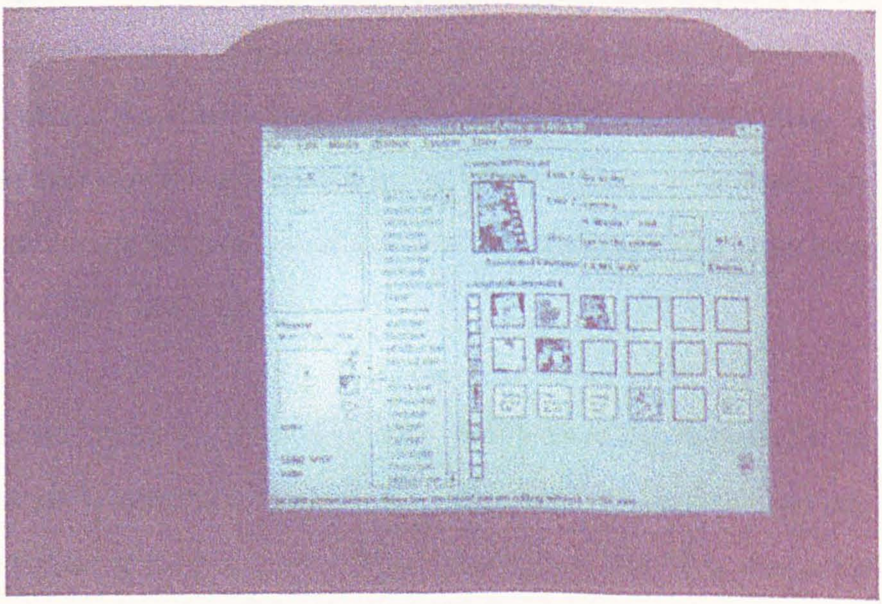
Although Easy Speaker was able to be utilised by moderately Learning Disabled adults there is potential to carry out trials with groups who are more severely handicapped. Obviously to verify the correctness of the approach severely handicapped individuals should be able to use Easy Speaker as their primary communication channel. This was not possible within the current study as a longer time frame would need to be studied; and during this study more able users were chosen as they could give some feedback as to their opinions toward Easy Speaker. More crucially initial research, such as that discussed within this thesis, cannot be carried out with such vulnerable groups without first verifying the approach and establishing an academic track record. Now that the approach has been validated it is hoped that professionals and caregivers in the area will take-up the work with more severely handicapped individuals as part of their working practice. Hardware cannot simply be lent short-term to such groups and then withdrawn when they become dependent on it simply for research purposes. Other parallel studies using Easy Speaker on a one to one basis with a variety of users suffering assorted handicaps have proved encouraging however. For example, Rostron, Plant and Hermann (1994) used Easy Speaker successfully with a severely Learning Disabled client who could potentially have used it as his primary means of communication had smaller more portable equipment been available at the time.

With improvements in both the range and quality of media available for linking to symbols further research as to its usefulness and impact needs to be carried out. In terms of improvements for maintainers the new object orientated maintainer interface needs to be evaluated in the field. Hopefully this will be assessed with both new maintainers and with those who used previous versions of Easy Speaker. Again it is hoped that professionals and caregivers in the field will comment on such improvements and provide feedback.

Future plans for research include work with non-vocal children and adolescents, with patients with degenerative diseases such as MS (Multiple Sclerosis) and MND (Motor Neurone Disease), and with laryngectomy patients. Hopefully this research will be carried out at hospitals in Hull and Sheffield in the UK from 1996 onwards using small highly portable devices such as the Fujitsu Stylistic 500 shown below running Easy Speaker Pro.

Fig 11.15, The stylistic 500 running Easy Speaker Pro





Currently collaborative work with the Voice Attitudes and Emotions in Speech Synthesis project (VAESS) based in Sheffield as part of the European TIDE program has begun. Partners include :-

- Sheffield University - Dept of Human Communication Science
- Stockholm - KTH
- Aalborg University - Centre for Personkommunikation (CPK)
- Madrid University - Speech Technology Group (GTH)
- Telia AB - Swedish Telecom
- Infovox
- BiDesign Ltd

This has meant that the Easy Speaker Pro software has been adapted to use both digitised speech and advanced synthesis systems such as Infovox OVE. As part of the VAESS project is to develop a suitable PC compatible platform in an A5 sized package weighing around a pound thus Easy Speaker Pro will run without modification using the touch screen input the device offers. In this role Easy Speaker will be required to run various symbol based systems in numerous languages. Focus in this area however has been on the quality of the speech synthesis in the main and the construction of a new rule based system.

With advances in available technology it is hoped to publicise and distribute Easy Speaker Pro by means of the Internet. We are currently constructing a World Wide Web site devoted to this purpose. Hopefully, by making Easy Speaker Pro available world wide, this potentially means that individuals who might benefit most may make use of it. In addition to which such distribution and use can only further some of the research goals outlined.

SUMMARY

Easy Speaker has been developed based on a three stage model which focuses on theoretical, developmental, and evaluative processes. All three were completed within a framework of existing, new and cutting-edge technologies.

The theoretical concentrated on the abilities and handicaps of potential users, and the need to provide a low cognitive demand ACD. The importance of a low demand, high synergy interface provided by a hypermedia metaphor was stressed in relation to cognitive modelling. The importance of concrete symbology was also outlined based on previous research. A model of how users interacted with Easy Speaker was provided by the SSOA model as applied to direct manipulation. Finally, the contentious issue of whether to use digitised or synthesised speech was addressed from both theoretical and experimental standpoints. Experimentally digitised speech incurred significantly less cognitive load.

The development of Easy Speaker was outlined from its inception through to the current version, and beyond. The importance of good design from both users and maintainers perspective's were stressed. In addition the contribution made by available hardware and software was considered briefly. On-line help and the need for enhanced error messages was highlighted. The ability of Easy Speaker to monitor users interactions was considered crucial and was a cornerstone of development. Maintainer's views were assessed by the QUIS. Overall maintainers QUIS responses were significantly favourable.

The evaluative stage focused on the levels of competence shown by real users. This focused on the completion of set tasks known as ESAT tasks. These measured constructional speed and efficiency. Results showed that users, despite their handicaps, could learn to operate Easy Speaker relatively rapidly, and that their performance soon reached optimum levels. Some even approached the level of skill shown by experienced unimpaired individuals. Crucially all used Easy Speaker productively and could easily generate multi-symbol utterances that were salient and semantically correct. Over the longer term off-device improvements in language and cognition were shown to have taken place as rated by two of each users immediate caregivers on the ESPM questionnaire. That is, when measured in adaptive behavioural terms as opposed to formal test measures. This suggested that not only could Easy Speaker be used as a primary or secondary communication channel, but also as a constant therapist. The aim being to improve residual linguistic and cognitive abilities within the user.

As suggested by the three stage model, development is a never ending cycle, fed by user needs, maintainer demands, and increasingly capable technologies. The latest version of Easy Speaker was outlined with the aim of addressing many of the shortcomings of the previous version. The new features it incorporates were listed along with possible implications for users and maintainers alike. The state of hardware was considered in relation to the next generation of software based ACD. An ideal hardware platform for the new version of Easy Speaker was also proposed. The change to software, over hardware based ACD's is seen as a welcome paradigm shift.

Other corroborative research which has been undertaken using Easy Speaker as its base was outlined, and where supportive of the current study parallels drawn.

Future directions for research were outlined, as was the need to achieve world wide publicity and distribution via the Internet and World Wide Web site. A list of useful addresses is provided in appendix A, detailing suppliers of commercial ACD's along with our Internet address.

BIBLIOGRAPHY

Abbeduto L, Furman L & Davies B, 1989, "Relation between the receptive language and mental age of persons with mental retardation", American Journal on Mental Retardation, 93, 5, 535-543, American Association on Mental Retardation

Abrahamsen AA, Romski MA & Sevcik RA, 1989, "Concomitants of success in acquiring an augmentative communication system: Changes in attention, communication and sociability", American Journal on Mental Retardation, 93, 5, 475-496, American Association on Mental Retardation

Aitkenhead AM & Slack JM, 1994, "Issues in cognitive modelling", Lawrence Earlbaum Associates

Alm N., 1988, "Towards a communication aid for severely physically disabled non-speaking people", PhD Thesis, Dept. of Mathematics and Computer Science, University of Dundee, Dundee, Scotland

Alm N, Arnott JL & Newell AF, 1992, "Prediction and conversational momentum in an augmentative communication system", Communications of the ACM, Vol 35, No 5

Aman MG, 1991, "Assessing psychopathology and behaviour problems in persons with mental retardation", US Department of Health and Human Services

Ammon PR & Ammon MS, 1971, "Effects of training black pre-school children in vocabulary vs sentence construction", Journal of Educational Psychology, 62, 421-426

Anderson JR, 1983, "The architecture of cognition", Harvard: Harvard University Press

Anderson M, 1988, "Inspection time, information processing and the development of intelligence", British Journal of Developmental Psychology, 6, 43-57

Anderson M, 1990, "Intelligence and the development of intellect", Spectrum, British Science News, 222, 2-5

Anderson M, 1992, "Intelligence and development: A cognitive theory", Blackwell Publishers

Arnheim R, 1972, "Visual thinking", University of California Press, Berkeley, CA

Baecker R, Small I & Mander R, 1991, "Bringing icons to life", Proc CHI'91: Human Factors in Computer Systems, ACM, New York, 1-6

Baird P, 1989, "Introduction", Hypermedia, Vol 1, No 1, 1-2

Barber P, 1990, "Applied cognitive psychology: An information-processing framework", Routledge

Bard EG, Shillcock RC & Altmann GTM, 1988, "The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context", Perception and Psychophysics, 29, 191-211

Barlow DH & Hersen M, 1984, "Single case experimental designs: Strategies for studying behaviour change", Pergamon Press Inc

Bateman B, 1972, "Learning disabilities - Yesterday, today and tomorrow", Exceptional Children, 31, 167-177

Berry DC & Dienes Z, 1991, "The relationship between implicit memory and implicit learning", British Journal of Psychology, 82, 359-373, The British Psychological Society

Bess F & Humes LE, 1995, "Audiology: The fundamentals (2^{nd} Ed)", Baltimore: Williams & Wilkins

Beukelman DR, 1988, "The use of microcomputer technology with persons unable to speak: An overview", Computers in Human Behaviour, Vol 4, pp 355-366, 1988

Beukelman DR & Ansel BM, 1995, "Research priorities in Augmentative and Alternative Communication", Augmentative and Alternative Communication, 11, 131-134

Blackwell CL, Hulbert CM, Bell J, Elston L, Morgan W, Robertshaw BA & Thomas C, 1989, "A survey of the communication abilities of people with a mental handicap", The British Journal of Mental Subnormality, XXXV, 1, 68

Borenstein NS, 1985, "The design and evaluation of On-Line Help systems", PhD dissertation, Dept of Computer Science, Carnegie-Mellon University, Pittsburgh, PA

Boyle C, Hor Teh S & Williams C, 1990, "An empirical evaluation of hypertext interfaces", Hypermedia, Vol 2, No 3, 235-247

Bousfield WA, 1953, "The occurrence of clustering in recall of randomly arranged associates", Journal of General Psychology, 49, 229-240

Bower GH & Winzenz D, 1969, "Group structure, coding and memory for digit series", Journal of Experimental Psychology Monograph, 80, 2, 2, 1-17

Brewer N & Nettelbeck T, 1979, "Speed and accuracy in the choice reaction time of mildly retarded adults", American Journal of Mental Deficiency, 84, 55-61

Brewer M, Kirby NH & Meeuwissen R, 1980, "Control of associated movement as a factor influencing reaction times of mentally retarded adults", American Journal of Mental Deficiency, 85, 281-285

Brown C, 1992, "Assistive technology for people with disabilities", Communications of the ACM, Vol 35, No 5

Brown H, 1991, "Hypermedia/hypertext and object orientated databases", Chapman & Hall

Bush V, 1945, "As we may think", Atlantic Monthly, 76, 1, July 1945, 101-108

Butterfield EC & Belmont JM, 1977, "Assessing and improving cognitive functions of mentally retarded people", In I Bialer and M Sternlicht (eds), The psychology of mental retardation: Issues and approaches, New York: Psychological dimensions

Butterfield EC, Wambold C & Belmont JM, 1973, "On theory and practice improving short-term memory", American Journal of Mental Deficiency, 77, 654-669

Carlson PA, 1990, "The rhetoric of hypertext", Hypermedia, Vol2, No 2, 109-131

Case R, 1985, "Intellectual development: Birth to adulthood", London: Academic Press

Chen SHA & Bernard-Opitz V, 1993, "Comparison of personal and computer assisted instruction for children with Autism", Mental Retardation, 368-376, American Association on Mental Retardation

Churcher PR, 1990, "A common notation for knowledge representation, cognitive models, learning and hypertext", Hypermedia,1,3

Chute DL, Conn G, Dipasquale MC & Hoag M,1988, "ProsthesisWare: A new class of software supporting the activities of daily living", Neuropsychology, 2, 41-57

Clark CR, 1981, "Learning words using traditional orthography and the symbols of Rebus, Bliss and Carrier", Journal of Speech and Hearing Disorders, 46, 191-196

Clark JE, 1983, "Intelligibility comparisons for two synthetic and one natural speech source", Journal of Phonetics, 11, 37-49

Clements J (ed), 1992, "Severe learning disability and psychological handicap", John Wiley & Sons

Cohill AM & Williges RC, 1982, "Computer-augmented retrieval of HELP information for novice users", Proc Human Factors Society, 26th Annual Meeting, 79-82

Collins A & Quillian MR, 1969, "Retrieval time from semantic memory", Journal of Verbal Learning and Verbal Behaviour, 8, 240-247

Conklin J, 1987, "Hypertext: An introduction and summary", IEEE Computer, Sept, 17-41, Elm and Woods,1985

Conrad C, 1972, "Cognitive economy in semantic memory", Journal of Experimental Psychology, 92, 148-154

Cooper J, Moodley M & Reynell J, 1986, "Helping language development: A developmental programme for children with early language handicaps", Edward Arnold

Cragg R & Garvey K, 1990, "What's on ?: A comprehensive menu of ordinary living activities for adults", R Cragg, 302 Station Rd, Kings Heath, Birmingham, B14 7TF

Crystal D, 1980, "Introduction to language pathology", Edward Arnold

Crystal D, Fletcher P & Garman M, 1976, "The grammatical analysis of language disability: A procedure for assessment and remediation", Edward Arnold

Darwin CJ, 1990, "Speech perception" in MW Eysenck (ed), The Blackwell dictionary of cognitive psychology, Oxford: Blackwell

Davison GC & Neale JM, 1990, "Abnormal Psychology (5th ed)", John Wiley & Sons

de Volder H, 1992, "Development of a hypertext-based speech aid", Unpublished MSc dissertation for the degree of Master of Information and Manufacturing, University of Hull, UK

Demasco PW & McCoy KF, 1992, "Generating text from compressed input: An intelligent interface for people with severe motor impairments", Communications of the ACM, Vol 35, No 5

Detterman KD, Mayer JD, Caruso DR, Legree PJ, Conners FA & Taylor R, 1992, "Assessment of basic cognitive abilities in relation to cognitive deficits", American Journal on Mental Retardation, 97, 3, 251-286

Dillon A, 1991, "Human factors issues in the design of hypermedia interfaces", Hypermedia/Hypertext and Object-Oriented Databases, ed Heather Brown, Ch 6, Chapman and Hall

Dodd B & Campbell R, 1986, "Hearing by eye: The psychology of lip reading", London: Lawrence Erlbaum

Doland VM, 1989, "Hypermedia as an interpretative act", Hypermedia, Vol 1, No 1, 6-19

Duchastel PC, 1990, "Examining cognitive processing in hypermedia usage, Hypermedia", Vol 2, No 3, 221-233

Ducker PC and Remmington B, 1991, "Manual sign-based communication for individuals with severe or profound mental handicap" in The challenge of severe mental handicap: A behaviour analytic approach,167-187, John Wiley & Sons

Duncan EB, 1989, "Structuring knowledge bases for designers of learning materials", Hypermedia, Vol 1, No 1, 20-33

Ellis A & Beattie G, 1993, "The psychology of language and communication", Lawrence Earlbaum Associates

Ellis AW & Young AW, 1988, "Human cognitive neuropsychology", Hove and London: Lawrence Erlbaum Associates

Ellis NR, 1970, "Memory processes in retardates and normals" in NR Ellis (Ed) International review of research in mental retardation, 4, 1-32, New York: Academic Press

Ellis NR, Woodley-Zanthos P & Dulaney CL, 1989, "Memory for spatial location in children, adults and mentally retarded persons", American Journal on Mental Retardation, 93, 5, 521-527, American Association on Mental Retardation

Elm WC & Woods DD, 1985, "Getting lost: A case study in interface design", Proceedings of the Human Factors Society 29^{th} Annual Meeting, 1985, 927-931

Engelbart D, 1984, "Authorship provisions in AUGMENT", Proc. IEEE CompCon Conference, 1984, 465-472

Eysenck MW & Keane MT, 1992, "Cognitive psychology: A students handbook", Lawrence Erlbaum Associates

Fay E, 1993, "Multimedia on the desktop", Computer Shopper, May 1993

Feagans L, 1983, "Discourse processes in disabled children" in JD McKinney & L Feagans (Eds) Current topics in learning disabilities, Norwood, NJ: Abex

Federico PA, 1992, "Assessing semantic knowledge using computer-based and paper-based media", Computers in Human Behaviour, Vol 8, 169-181, Pergamon Press Ltd

Felce D, 1994, "Facilitated Communication: Results from a number of recently published evaluations", British Journal of Learning Disabilities, 22, 122-126

Fitts PM & Posner MI, 1967, "Human performance", Brookes Cole, Balmont CA

Foulds R, 1980, "Communication rates for non-speech expression as a function of manual tasks and linguistic constraints", Proceedings of International Conference on Rehabilitation Engineering, Toronto

Frauenfelder UH & Tyler LK, 1987, "The process of spoken word recognition: An introduction", Cognition, 25, 1-20

Fucci D, Reynolds ME, Bettagere R & Gonzales MD, 1995, "Synthetic speech intelligibility under several experimental conditions", Augmentative and Alternative Communication, 11, 113-117

Garrett MF, 1976, "Syntactic processes in sentence production" in RJ Wales & E Walker (eds), New approaches to language mechanisms, Amsterdam: North Holland

Garrett MF, 1984, "The organisation of processing structures for language production: Applications to aphasic speech", In D Caplan, AR Lecours & A Smith (Eds), Biological perspectives on language, Cambridge, Mass: MIT Press

Gibbs DP & Cooper EB,1989, "Prevalence of communication disorders in students with Learning Disabilities", Journal of Learning Disabilities, 22, 1, 60-63

Glinert EP & York BW, 1992, "Computers and people with disabilities", Communications of the ACM, Vol 35, No 5

Goodman J & Remmington B, 1991, "Teaching communicative signing: Labelling, requesting and transfer of function" in The challenge of severe mental handicap, 215-234, John Wiley & Sons

Goswami U & Bryant P, 1990, "Phonological skills and learning to read", Hove: Erlbaum

Greenfield DB & Oller DK, 1989, "Design and methodological requirements for studying the role of language acquisition in cognitive/social development", American Journal on Mental Retardation, 93, 5, 497-502, American Association on Mental Retardation

Groff P, 1990, "An analysis of the debate: Teaching reading without conveying phonic information", Interchange, 21, 4:1-14

Hampton JA, 1979, "Polymorphous concepts in semantic memory", Journal of Verbal Learning and Verbal Behaviour, 18, 441-461

Hansen WJ, 1971, "User engineering principles for interactive systems", Proc Fall Joint Computer Conference 39, AFIPS Press, Montvale, NJ, 523-532

Hick WG, 1952, "On the rate of gain of information", Quarterly Journal of Experimental Psychology, 4, 11-26

Higginbotham DJ & Baird E, 1995, "Analysis of listeners' summaries of synthesised speech passages", Augmentative and Alternative Communication, 11, 101-112

Hurlbut BI, Iwata BA & Green JD, 1982, "Nonvocal language acquisition adolescents with severe physical disabilities: Bliss symbols versus iconic stimulus formats", Journal of Applied Behaviour Analysis, 15, 241-258

Hutt E, 1986, "Teaching language-disordered children: A structured curriculum", Edward Arnold

Hyman R, 1953, "Stimulus information as a determinant of reaction time", Journal of Experimental Psychology, 45, 188-96

Jensen AR, 1979, "*g*: Outmoded theory or unconquered frontier ?", Creative Science and Technology, 1, 799-803

Jensen AR, 1980, "Chronometric analysis of mental ability", Journal of Social and Biological Structures, 3, 181-224

Jensen AR, 1982, "Reaction time and psychometric journal" in HJ Eysenck (Ed), A model for intelligence (93-94), Berlin: Springer Verlag

Jensen AR, 1987b, "Process differences and individual differences in some cognitive tasks", Intelligence, 11, 107-36

Johnson P, 1992, "Human computer interaction: Psychology, task analysis and software engineering", McGraw-Hill Book Co

Jonassen DH, 1989, "Semantic networking approaches to structuring hypertext", Paper presented at Hypertext 11 York, West Yorkshire, June 29-30 1989

Kail R, 1988, "Reply to Stigler, Nusbaum and Chalip", Child Development, 59, 1154-7

Kail R, 1992, "General slowing of information-processing by persons with mental retardation", American Journal on Mental Retardation, 97, 3, 333-341

Katz JJ & Fodor JA, 1963, "The structure of a semantic theory", Language, 39, 170-210

Kintsch W, 1970, "Models for free recall and recognition" in DA Norman Models of human memory, London: Academic Press

Kirby NH, Nettelbeck T & Western P, 1982, "Locating information processing deficits of mildly mentally retarded young adults", American Journal of Mental Deficiency, 97, 338-343

Klahr D, 1989, "Cognition and instruction", Hillsdale, New Jersey: Lawrence Erlbaum

Klausmeir HJ & Harris CW, 1966, "Analysis of concept learning", Academic Press

Kukich K, 1992, "Spelling correction for the telecommunications network for the deaf", Communications of the ACM, Vol 35, No 5

Lally M & Nettlebeck T, 1977, "Intelligence, reaction time and inspection time", American Journal of Mental Deficiency, 82, 273-81

Leonard LB, Chapman K, Rowan LE & Weiss AL, 1983, "Three hypotheses concerning young children's imitations of lexical items", Developmental Psychology, 19, 591-601

Lloyd LL, 1989, Augmentative and alternative communication, American Journal on Mental Retardation, 93, 5, 500-503, American Association on Mental Retardation

Mackay HA, 1991, "Stimulus equivalence: Implications for the development of adaptive behaviour" in The challenge of severe mental handicap, Edited by Remmington B, 235-259, John Wiley & Sons

Magers CS, 1983, "An experimental evaluation of on-line HELP for non-programmers", Proc CHI'83 Conference: Human Factors in Computing Systems, ACM, New York, 277-281

Mansell JL, 1993, "Services for people with learning disabilities and challenging behaviour or mental health needs: Report of a project group", London: HMSO

Marge M, 1972, "The general problem of language disabilities in children" in JV Irwin & M Marge (Eds), Principles of childhood language disabilities, Englewood Cliffs, NJ: Prentice-Hall

Marslen-Wilson W & Tyler LK, 1980, "The temporal structure of spoken language understanding", Cognition, 8,1-71

Mattingly IG & Liberman AM, 1988, "Specialised perceiving systems for speech and other biological significant sounds" in GW Edelman, WE Gall & WM Cowan (eds), Functions of the auditory system, New York: Wiley

McGarry MS & West SG, 1975, "Stigma among the stigmatised: Resident mobility, communication ability and physical appearance as predictors of staff - resident interaction", Journal of Abnormal Psychology, 84, 4, 399-405

McGregor A and Alm N, 1992, "Thoughts of a non-speaking member of AAC research team", Paper delivered to the Sixth International Conference, International Society for Augmentative and Alternative Communication, Philadelphia

McKnight C, Dillon A & Richardson J, 1989, "Problems in hyperland ? A human factors perspective", Hypertext, Vol1, No 3, 167-178

McNamara TP & Sternberg RJ, 1983, "Mental models of word meaning", Journal of Verbal Learning and Verbal Behaviour

Merrill EC & Jackson TS, 1992, "Sentence processing by adolescents with and without mental retardation", American Journal on Mental Retardation, 97, 3, 342-350

Michon JA, 1972, "Multidimensional and hierarchical analysis of progress in learning" in LW Gregg (ed) Cognition in learning and memory, New York: Wiley

Miller E, 1993, "Dissociating single cases in neuropsychology", British Journal of Clinical Psychology, 32, 155-167, The British Psychological Society

Mirrenda P & Beukelman DR, 1987, "A comparison of speech synthesis intelligibility with listeners from three age groups", Augmentative and Alternative Communication, 3, 120-128

Mirenda P & Beukelman DR, 1990, "A comparison of intelligibility among natural speech and seven speech synthesisers with listeners from three age groups", Augmentative and Alternative Communication, 6, 61-68

Mirenda P & Locke PA, 1989, "A comparison of symbol transparency in nonspeaking persons with intellectual disabilities", Journal of Speech and Hearing Disorders, 54, 131-140

Mitchell PR & Adkins CP, 1989, "A comparison of single word intelligibility of two voice output communication aids", Augmentative and Alternative Communication, 5, 84-88

Monk A, 1984, "Fundamental of human-computer interaction", Academic Press: London

Moody G,1992, "Face to face", Windows User, December 1992

Muller DJ, Munro SM & Code C, 1981, "Language assessment for remediation", Croom Helm

Nelson KE, 1989, "Indirect aspects of communicative gains: Toward fuller accounts of sequence and process", American Journal of Mental Retardation, 93, 5, 503-505, American Association on Mental Retardation

Nettlebeck T, 1985, "Inspection time and mild mental retardation", International Review of Research in Mental Retardation, 13

Newall A, 1991, "Assisting interaction with technology: Research and practice", British Journal of Disorders of Communication, 26, 1-10

Newell AF & Alm N, 1994, "Developing AAC technologies: a personal story and philosophy", European Journal of Disorders of Communication, 29, 399-411

Newman H, 1982, "The sounds of silence in communicative encounters", Commun. Quart., 30, 2

Norman DA, 1975, "Cognitive organisation and learning" in PMA Rabbit & S Domic Attention and Performance, London: Academic Press

Norman DA, Rumelhart DE & The LNR Research Group, 1975, "Explorations in cognition", San Francisco: Freeman

Norman K, 1991, "The psychology of menu selection: Designing cognitive control at the Human/Computer Interface", Ablex, Norwood, NJ

Oren T, Salomon G, Kreitman K & Don A, 1990, "Guides: Characterising the interface" in Laural B (Ed) The art of Human-Computer Interface design, Addison Wesley, Reading, MA, 367-381,

Panyan M, 1984, "Computer technology for autistic students", Journal of Development Disorders, 14, 375-382

Papert S, 1980, "Mindstroms: Children, computers and powerful ideas", Basic books, New York

Pausch R & Williams RD, 1992, "Giving CANDY to children: User-tailored gesture input driving an articulator based speech synthesiser", Communications of the ACM, Vol 35, No 5

Plant RR, 1991, "An evaluation of access systems for speech prostheses", Unpublished second year degree project

Plant RR, 1992, "An evaluation of constructional methods used in the building of English sentences using a hypermedia Graphical User Interface", Unpublished final year research project

Plant RR & Rostron AB, 1996, "Assistive technology - whose needs are greatest - users or care-givers ? A note on software design", Information Technology and Disabilities Journal (RIT), ITD, Vol 3, No 1, Feb 1996, ISSN 1073-5127

Porter DB, 1991, "Computer games and cognitive processes: Two tasks, too modes, too much ?", British Journal of Psychology, 82, 343-357,The British Psychological Society

Pratt RL, 1987, "Quantifying the performance of text-to-speech synthesisers", Speech Technology, 3, 54-64

Rabbitt PMA & Maylor EA, 1991, "Investigating models of human performance", British Journal of Psychology, 82, 259-290,The British Psychological Society

Rada R & Barlow J, 1988, "Expert systems in hypertext", The Knowledge Engineering Review, 3(4), 1988

Reichle J, Sigafoos J and Remmingtom B, 1991, "Beginning an augmentative communication system with individuals who have severe disabilities" in The challenge of severe mental handicap edited by Bob Remmington, John Wiley & Sons Ltd

Remmington B, 1991, "The challenge of severe mental handicap: A behaviour analytic approach", John Wiley & Sons

Repp AC & Karsh K, 1991, "The task demonstration model: A program for teaching persons with severe disabilities" in The challenge of severe mental handicap, Edited by Remmington B, 263-283, John Wiley & Sons

Rizk A, Streitz N & André J (Eds), 1990, "Hypertext: Concepts, systems and applications: Proceedings of the European conference on hypertext", INRIA, France, November 1990, Cambridge University Press

Rosch E, 1972, "Natural categories", Cognitive Psychology, 4, 328-350

Rostron AB & Plant RR, 1992, "Finding a voice on disk", British Journal of Special Education, Vol 19, 4

Rostron AB, Plant RR & Hermann C, 1994, "Hypermedia for the learning disabled: Expensive luxury or useful tool", Computers in Education, 22, 3, 215-223

Rostron AB & Plant RR, 1994, "Some applications of a hypermedia-based speech synthesiser for the handicapped", presented at Computers in Psychology 94, University of York, 21st - 23rd Sept 1994

Rostron A & Kinsella P, 1995, "Foreign language pronunciation using CALL: Some experimental evidence", ReCALL, Newsletter of the CTI Centre for Modem Languages, The University of Hull

Rostron A & Gimson, 1995, "Communication aids for the hearing impaired: Some recent experiences with a multimedia system", British Journal of the Association of Teachers of the Deaf, 19, 4, 111-117

Rostron A, Plant RR & Ward, 1996, "Computerised augmentative communication devices for people with dysphasia: Design and evaluation", European Journal of Disorders of Communication, March 1996 (in press)

Rostron AB & Plant RR, 1996, "Assistive Technology: whose needs are greatest - users or care-givers ?", presented at Computers in Psychology 96, University of York, 25th - 27th March 1996

Rumelhart DE, McClelland JL & The PDP Research group, 1986, "Parallel distributed processing: Explorations in the microstructure of cognition", Vol 1:"Foundations", Cambridge Mass.: Bradford Books, MIT Press

Sawyer DJ & Fox BJ, 1991, "Phonological awareness in reading", New York: springer-Verlag

Schank RC, 1975, "Conceptual information processing", Amsterdam: North-Holland

Scherer K, 1979, "Personality markers in speech" in Scherer K (Ed) Social markers in speech, Cambridge University Press, London

Shannon CE & Weaver W, 1949, "The mathematical theory of communication", Urbana, Il: University of Illinois Press

Shiffrin RM & Dunmis ST, 1981, "The development of automatism" in JR Anderson (Ed), Cognitive skill and their acquisition, 111-140, Hillsdale, NJ: Erlbaum

Shneiderman B, 1980, "Software psychology: Human factors in computer and information systems", Little, Brown, Boston, MA

Shneiderman B, 1983, "Direct manipulation: A step beyond programming languages", IEEE Computer 16, 8, 57-69

Shneiderman B, 1992, "Designing the user interface: Strategies for effective human-computer interaction", Addison Wesley

Shum S, 1990, "Real and virtual spaces: Mapping from spatial cognition to hypertext", Hypermedia, Vol 2, No 2, 133-158

Snider VE, 1992, "Learning styles and learning to read: A critique", Remedial and Special Education, 13, 1:6-18

Songco DC, Allen SI, Plexico PS & Marford RA, 1980, "How computers talk to the blind", IEEE Spectrum, May 1980, 34-38

Sorrentino RM & Higgins ET, 1991, "Motivation and cognition: Warming up to synergism, Handbook of motivation and cognition: Foundations of social behaviour", 3-19, John Wiley & Sons

Srull TK & Wyer RS (Jnr), 1991, "The role of chronic and temporary goals in social information processing" in the Handbook of motivation and cognition: Foundations of social behaviour, 503-549, John Wiley & Sons

Stephens M, 1993, "The face at the window", Windows User, January 1993

Strathmeyer CR, 1990, "Voice in computing: An overview of available technologies", IEEE Computer, 23, 8, 10-16

Strolley BT, Rabbitt PMA & Moore BJ, 1990, "Speed and concentration in healthy old age", Paper given at the Third Cognitive Ageing Conference, Atlanta, April

Symth MM, Morris PE, Levy P & Ellis AW, 1993, "Cognition in Action", Lawrence Earlbaum Associates

Tomporowski Pd, Hayden Am & Applegate B, 1990, "Effects of background event rate on sustained attention of mentally retarded and nonretarded adults", American Journal on Mental Retardation, 94, 5, 499-508

Tulving E, 1972, "Episodic and semantic memory" in R Tulving and W Donaldson (Eds) Organisation of memory, London: Academic Press

Tunmer WE, 1990, "The psychology of learning to read: current rends in theory and research", Education Research and Perspectives, 17, 1:34-49

Vaina L & Hintikka J (Eds), 1984, "Cognitive constraints on communication: Representations and processes", D Reidel Publishing Co

Van Dam A, 1988, "Hypertext 87: Keynote address", Communications of the ACM, 31, 7, (July 88), 887-895

Van Oosterom J & Devereux K, 1992, "The rebus glossary", Learning Development Aids

Vernon PA, 1987, "Speed of information-processing and intelligence", Norwood, NJ: Albex

Verplank WL, 1988, "Graphic challenges in designing object-oriented user interfaces" in Helander (ed), Handbook of Human Computer Interaction, Elsevier Science Publishers, Amsterdam

Vygotsky IS, 1978, "Mind in society: The development of higher psychological processes", Edited and translated by M Cole, V Joun-Steiner, S Scribner & E Souberman, Cambridge Mass: Harvard University Press

Waterworth JA & Chingnell MH, 1989, "A manifesto for hypermedia usability research", Hypermedia, Vol 1, No 3, 205-234

Westwood P, 1993, "Common-sense methods for children with special needs", Second Edition, Routledge

Whittington KR, 1989, "Speech processing linking concept to application", Higher Education - Enterprise Link Series, Hobsons Publishing PLC

Wiersman UJ, 1992, "The effects of extrinsic rewards in intrinsic motivation: A meta analysis", British Journal of Occupational and Organisational Psychology, 65, 101-114, The British Psychological Society

Wilhite SC, 1991, "Evidence of a negative environmental reinstatement effect", British Journal of Psychology, 82, 325-342, The British Psychological Society

Wright P, 1989, "Interface alternatives for hypertexts", Hypermedia, Vol 1, No 2, 146-166

Yankelovich N, Meyrowitz & Van Dam A, 1985, "Reading and writing the electronic book", IEEE Computer, 18, 10, 15-30

Yerkes RM & Dodson JD, 1908, "The relation of strength of stimulus to rapidity of habit formation", Journal of Comparitive and Neurological Psychology, 18, 459-482

Yule G, 1986, "The study of language", Cambridge University Press

Wade MG, Newell KM & Wallace SA, 1978, "Decision time and movement time as a function of response complexity in retarded persons", American Journal of Mental Deficiency, 83, 135-144

Walberg HJ, 1988, "Synthesis of research on time and learning", Educational Leadership 45, 6, 76-85

Waltz DL & Pollack JB, 1985, "Massively parallel parsing: A strongly interactive model of natural language interpretation", Cognitive Science, 9, 51-74

Wurman RS, 1989, "Information anxiety", Doubleday, New York

APPENDIX A: Useful contact addresses

Currently a demonstration version of Easy Speaker is available to download free on a trial basis via a World Wide Web site located at :-

http://www.hull.ac.uk/

From the main University page select the Psychology Dept's home page. From here there is access to a specific Communication Aids page with relevant software and contact addresses.

You can also contact me via this Internet site or I can be eMailed on :-

r.r.plant@psy.hull.ac.uk

These addresses will remain current for as long is as feasible.

Brilliant Computing/SEMERC, Box 142, Bradford, BD9 5NF

Cane & Able Ltd, Grindle Turret Main Road, Grindleford, Sheffield S30, South Yorkshire

Capital Electronic Developments Ltd, 590 Uxbridge Road, Hayes

The Computability Centre, PO Box 94, Warwick, CV34 5WS

Ezi-Line Ltd, 24 East Street, Wareham, BH 20

Hugh Steeper Ltd, Queen Mary's University Hospital, Roehampton Disability Centre, London, SW15 5PL

Liberator Ltd, Whitegates, Swinstead, Lincoln, NG33 4PD

Newman Tonks Group, Oxleason Road, East Moons Moat, reddich, B948 0RE

QED Ltd, Ability House, 242 Gosport Road, Fareham, PO16 0SS

QRO Systems, Valley View, Hadleigh Road, Ipswitch, Suffolk, IP2 0BT

Rainbow Rehab, 134 Purewell Road, Cristchurch, BH23 1EU

Raymar, Unit 1, Fairview Estate, Reading Road, Henly on Thames, RG9 1HE

ROMPA International, Goyt Side Road, Chesterfield, S40 2PH

Toby Churchill Ltd, 20 Panton St, Cambridge, CB2 1HP