Simulating naturalistic instruction: The case for a voice mediated interface for assistive technology for cognition

Dr Brian O’Neill
West of Scotland Mobility and Rehabilitation Centre
Southern General Hospital
1345 Govan Road
Glasgow, G51 4TF
UK
Tel: + 44 1412012391
Fax: + 44 1412012649
brian.oneill@sgh.scot.nhs.uk

Dr Alex Gillespie
Department of Psychology
University of Stirling
Scotland, FK9 4LA
UK

1 Author for correspondence
A variety of brain pathologies can result in difficulties performing complex behavioural sequences. Assistive technology for cognition (ATC) attempts support of complex sequences with the aim of reducing disability. Traditional ATCs are cognitively demanding to use and thus have had poor uptake. A more intuitive interface may allow ATCs to reach their potential. Insights from psychological science may be useful to technologists in this area. We propose that an auditory-verbal interface is more intuitive than a visual interface and reduces cognitive demands on users. Two experiments demonstrate a novel ATC, the General User Interface for Disorders of Execution (GUIDE). GUIDE is novel because it simulates normal conversational prompting to support task performance. GUIDE provides verbal prompts and questions and voice recognition allows the user to interact with GUIDE. Research with non-cognitively impaired participants and a single participant experiment with a person with vascular dementia provide support for using interactive auditory-verbal interfaces. Suggestions for the future development of auditory-verbal interfaces are discussed.

Keywords: Assistive technology, executive function, GUIDE, complex behaviour, verbal interface
Difficulties carrying out goal-directed behaviour lead to a high degree of disability for which there exist few treatment options. These difficulties are manifest in groups such as traumatic brain injury (Evans, 2003), schizophrenia (Semkovska, Bedard, Godbout, Limoge & Stip, 2003; Krabbendam, de Vugt, Derix & Jolles, 1999) learning disability (Cavalier & Ferretti, 1993) and the dementias, contributing to the high degree of personal care required by persons in these groups.

In samples of those with physical and cognitive disability, use of assistive technologies is associated with reduced need for personal assistance (Hoenig, Taylor & Sloan, 2003). However systems designed to support cognitive function can often be cognitively demanding, highlighting the need for systems to be useable by those with more severe cognitive disability (LoPresti, Mihailidis & Kirsch, 2004).

Traditional scheduling technology is complex to use. Diaries require users to recall where they are, remember to enter a prompt, remember to check the diary and monitor task completion. Recipes schedule sub-steps of a goal state. However they also require decoding of the written steps and information storage in working memory to underpin performance.

The application of digital technology to cognitive rehabilitation has also been limited by the cognitive demands of interfaces. Personal digital assistants (PDAs) and
palmtops are designed to extend the cognitive abilities of those without impairment and have a learning curve which places them beyond many with cognitive deficits. Designs of simplified interfaces has allowed use by those with learning disability to aid with time management (Davies, Stock, Wehmeyer 2002). Prospective memory aids have benefits in reducing omissions of to-be-performed behaviours. Specifically text prompts delivered to portable systems such as pagers and mobile phones increase the hit rate of target behaviours (Evans, 2003). These systems are useful for those with intact reading and direction of attention. For those with yet more severe cognitive difficulties we suggest that systems based on auditory-verbal interfaces may be more appropriate.

Executive Function, Language & Scaffolding

Developmental psychologists studying executive function emphasise its basis in language, and more particularly private speech (i.e., speech directed at self rather than an interlocutor). This developmental account was initially proposed by Vygotsky and Luria (1930/1994; Luria,1961). Although language evolved for communication, Vygotsky and Luria speculate that it has a secondary function enabling humans to talk themselves through complex tasks, thus facilitating the execution of complex purposive and planned behaviour.

Research on child development supports a relation between language and executive function (Hughes, 1996, 2002; Barkley, 1997; Jones, Rothbart & Posner, 2003; Zelazo, Muller, Frye & Marcovitch, 2003). Children engage in private speech when problem solving and this increases in probability with increasing task complexity.
(Berk & Garvin, 1984). There is also a prevalence of private speech problems amongst children who are at risk for attention and behaviour problems (Winsler, Diaz, Atencio, McCarthy & Chabay, 1985) and amongst children who do worse at tasks (Frauenglass & Diaz, 1985). In adults without cognitive impairment, introducing a secondary task which disrupts private speech disrupts performance on a planning task (Phillips, Wynn, Gilhooly, Della Sala & Logie, 1999).

Vygotsky and Luria proposed that the child initially has minimal executive control, and that executive control is that provided by other people, such as parents and carers. The verbal and visual actions of these carers guide the child’s behaviour, providing a ‘scaffold’ (Berk & Winsler, 1995; Gillespie, 2006; Zittoun, Gillespie, Cornish & Psaltis, 2007). There are activities that the child is unable to perform alone, but is able to perform with appropriate guidance. Such activities constitute the ‘zone of proximal development’ and scaffolding operates in this zone. The scaffold comprises mainly of verbal guidance. Verbal guidance is used to direct attention, set sub-goals, initiate monitoring, and correct errors. Children develop through this verbal scaffold by internalising the guidance so that they become able to verbally guide their own behaviour (Gillespie, 2007). The monitoring and regulation of behaviour initially occurs between the child and the carer, but with development becomes an intra-psychological function, namely, executive function.

The relevance of the concepts of scaffolding and the zone of proximal development for rehabilitation have been recognised (Stone, 1998; Young et al., 2002). In the same way that there are problems which children are unable to solve alone, but which they are able to solve with verbal guidance, so there people with a variety of cognitive
deficits who confront activities of daily living which they are unable to perform alone, but which they are able to perform with appropriate verbal guidance. We suggest that carers are often guiding patients at the limits of the patient's ability, providing cognitive support by using verbal prompts to scaffold patients’ executive function.

Initiation, problem-solving, generativity, planning, sequencing, organization, self-monitoring, error correction and behavioural inhibition are functions taken on by the carer if these are deficient in the cared-for. The basic cognitive abilities that the patient must possess in order to gain from this instruction are language comprehension, maintenance of a single command in mind and verbally mediated motor programming. Carers successfully support, or scaffold, individuals with executive dysfunction in carrying out activities of daily living by instruction (Gitlin, et al., 2002). In this sense, carers are often acting as highly refined ‘assistants for cognition’. Our aim has been to try and simulate the verbal scaffolding and guidance provided by carers.

Developing an Auditory-Verbal Interface

Assistive technologies for cognition, also known as cognitive prosthetics, have the potential to revolutionise the management of cognitive disabilities (Gregor & Newell, 2004). Prospective memory aids facilitate the performance of a behaviour at a time in the future which might otherwise be forgotten. Traditional prospective memory aids are commonly used and include paper notes, diaries, calendars, alarms and reminders (Evans, Wilson, Needham & Brentnall, 2003). These examples of traditional assistive technology have been augmented in recent years by digital assistive technology.
Personal digital organisers and voice recorders can now undertake the function of several of the traditional assistive technologies, such as temporal prompting and thus recall of the to-be-performed behaviour (Kapur, Gilisky & Wilson, 2004; Yasuda et al., 2002). Computer systems allow central storage of schedules to be delivered as text prompts at the point when a behaviour requires to be carried out. The effectiveness of pager prompts as used in the proprietary Neuropage system has been demonstrated to increase achievement of target behaviours (Wilson, Emslie, Quirk, & Evans, 2001). Similarly the MEMEX project, utilising text messaging to mobile phones, has demonstrated effectiveness in improving attendance at appointments and medication compliance (Pijnenborg, Withaar, Evans, van den Bosch & Brouwer, 2007).

LoPresti, Mihailidis and Kirsch (2004) suggest that ATCs have not been achieving their full potential in the main, due to the complexity of the ATC devices. Rather than reducing cognitive load, they often increase the cognitive burden by requiring users interact with complex and unfamiliar devices. Accordingly, these authors call for future ATC devices to be more sensitive in orienting to their cognitively impaired users (see also, Scherer, 2001).

Based on our review of the available ATCs we argue that they are biased toward visual interfaces. Visual interfaces are attention demanding. Attentional function predicts use of cognitive aids in sample of people with acquired brain injury (Evans, Wilson, Needham & Brentnall, 2003). Yet the main ATCs, such as Neuropage (Wilson, Emslie, Quirk, & Evans, 2001) and MemoJog (Inglis et al., 2002), require users to interact with the ATC device via a screen. Users usually receive information via the screen, and give feedback to the device via the screen. Visual interfaces are
good if the user’s visual system is free and the user is familiar with such interaction. However, in cases where a user is engaged in a task (such as dressing, food preparation, transfer or donning a limb) interacting with a computer screen entails a shift of visual attention and an interruption of the ongoing task, thus increasing the cognitive load of the task. Where users are unfamiliar with computers, the device will also be unintuitive contributing to low uptake.

We argue, on the basis of the literature on scaffolding, that an effective way forward is for ATCs to model the cognitive support provided by carers, or ‘assistants for cognition’. First, the cognitive scaffolding provided by carers is verbal, not visual, and thus does not lead the patients’ visual attention away from the task at hand. Second, given the evidence presented about the close relationship between executive function and language, we speculate that prompting in the verbal medium rather than the visual medium provides a more direct augmentation of executive function. Third, the verbal guidance provided by carers is in a familiar mode of interaction, namely, communicative interaction, and thus there is no learning curve. Finally, the scaffolding provided by carers is task-focused and tailored to the individual, that is to say each verbal prompt is contextually relevant to the immediate sub-goal that the user is engaged in. Written instructions, flow charts or diagrams, on the other hand, often present all the prompts and guidance at the same time, for example on the one sheet of paper or flow chart. Such information overload is avoided in the auditory medium because the linearity inherent in the auditory medium ensure that only one prompt is presented at a time.
We have developed an ATC that simulates the type of guidance provided by carers. The device is called GUIDE© - General User Interface for Disorders of Execution. The GUIDE© uses the most generalizable interface known for guiding users through complex tasks, namely verbal guidance. The GUIDE© prompts users, asks users questions and accepts verbal responses. The GUIDE© uses the verbal responses to direct the deployment of subsequent prompts and questions. Delivering prompts and questions to users verbally (rather than visually) entails less cognitive load, and does not require users to switch task-focus (such as shifting from task to viewing a screen or a card). The GUIDE© is designed to augment task focus, not to interrupt it. The GUIDE© does not disempower users, by undermining their agency, rather the GUIDE© empowers users. The GUIDE© prompts users with simple questions, requiring the user to engage in the task in order to answer. The range of answers that the GUIDE© accepts is deliberately limited, so as to reduce cognitive load. An assumption is that verbal responses are comparatively easy for users, requiring the least cognitive load and are similar to verbal interactions with carers.

The GUIDE© is in the tradition of COACH (Mihailidis, Barbenel & Fernie, 2004) in that it is focused upon a particular task and uses verbal prompts. The GUIDE© incorporates Mihailidis, Barbenel and Fernie’s (2004, p. 166) recommendations for constructing a verbal prompting system: prompts are provided at different levels of detail, prompts are tailored to the individual user, prompts only pertain to one action at a time, and prompts are phrased in a terminology that is familiar. The GUIDE© also goes beyond these recommendations by virtue of attempting explicitly to
simulate the scaffolding support provided by carers. The prompts used by the 
GUIDE© are modelled on the prompts provided by carers. Users interact with the 
GUIDE© in the same way that they do with carers, namely, verbal exchange. 
Specifically, this means that many of the prompts are actually questions, and these 
questions require a verbal response from the user. The questions engage the user and 
are intended to stimulate and augment, rather than command.

Research on a range of ATCs has tended to conclude that generic systems come into 
problems both due to the particularities of the task to be achieved and particularities 
of the patient (Cole, 1999; Mihailidis, Barbenel & Fernie, 2004). In order to avoid 
such problems protocols are carefully constructed to address the task to be achieved 
by the user.

Technologically, the GUIDE© has four components: hardware, voice recognition 
software, an action pathway used to ‘scaffold’ users behaviour, and the GUIDE© 
software program, running on a PC, that coordinates the parts. The action pathway 
module within the GUIDE© stores steps and checks based on observations of 
competent performance of the task. Verbal prompts are recorded and stored on the 
PC. Verbal responses are processed via voice recognition software.

In order to develop our argument that audio-verbal interfaces are particularly suited to 
ATCs which target behavioural sequences which require visual attention, we present 
data from two recent studies using GUIDE© (Dickie, Tyler, O’Neill & Gillespie, 
submitted; O’Neill & Gillespie, submitted). The first presents the results of a study of 
adults without cognitive impairment carrying out a novel task with a distracting
secondary task comparing GUIDE use with following written instructions. The second presents a case study of a man with vascular dementia carrying out a self-care behaviour which he failed to acquire in rehabilitation as usual.

Study 1: Using GUIDE© with non-impaired adults

*Design:* A between-participants design was used with participants randomly allocated to two groups. The task was to make an elaborate smoothie. One group received only written instructions while the other group received the same instructions via the GUIDE©. Periodic distraction was simulated in both conditions with a random number generation task.

*Participants:* 26 participant sample of convenience (mean age 26; range 19 to 43; 17 males, 9 females) randomly allocated to one of two groups. The Written Instruction Group comprised 10 male and 3 female. The GUIDE© Group comprised 7 male and 6 female. No significant age differences existed between the groups.

*Procedure:* Before beginning the task participants were trained in the use of GUIDE© for about 10 minutes. Training of the voice recognition software also occurred at this time. The five commands were repeated 15 times each to reach an adequate level of recognition. The task occurred in a kitchen with all equipment and ingredients present. During the smoothie making task, participants were asked to generate a random number between 34 and 43 every 30 seconds, throughout the experiment.
Scoring: Three indices of performance were independently rated: Percentage of steps completed successfully; frequency of hesitations, other than when reading or listening to prompts, of greater duration than 10 seconds; and mean time per step of the task. Inter-rater reliability, measured with Pearson’s correlation coefficient was high for number of correct steps (r=0.92, p<.001), for deviations (r=0.95, p<.001) and for hesitations (r=.73, p<.001).

Following completion of the task, participants completed a questionnaire designed to assess the participants’ feelings towards the method of instruction they experienced, whether written instructions or the Guide© device. This questionnaire was scored by counting the total number of positive and negative comments reported by the participant.

Results: Table 1 shows the group means for percentage of correct steps and incorrect steps, number of hesitations and average time per step.

For the number of steps completed correctly, the Guide© performed significantly better than the written instructions (F(1) = 6.375, p<0.05). The number of hesitations show by users of the Guide© was significantly fewer than those shown by the group following written instructions (F(1) = 12.084, p<0.05). The average time per step was not significantly different, however there was a trend towards the Guide© steps being completed more quickly than those in the written instruction group (F(1) = 1.326, p>0.05). The mean number of positive comments given in the questionnaires was 2.5
for the Guide© (SD 1.31) and 1.0 for the written instructions (SD 1.04). This was significantly different (F(1) = 9.581, p<0.05).

**Study 2: Using GUIDE© with a cognitively impaired user**

*Participant:* A 67 year old man (BB) who underwent bilateral transtibial amputations due to peripheral arterial disease and comorbid diabetes a year prior to this intervention. Following amputation he had received out-patient rehabilitation in a specialist centre for post-amputation rehabilitation three times weekly for 8 weeks. This consisted of instruction from a physiotherapist on the correct donning of his prosthetic limbs and mobilization using a variety of supportive walking aids. He consistently made errors in donning his prosthetic limbs and omitted critical safety checks. He was thus advised by his physiotherapists that he was not to don his prostheses without supervision. A standard neuropsychological assessment (RANDOLPH, 1998) demonstrated impairments across several cognitive domains with visuospatial function, attention and delayed memory lying in the extremely low range.

Provision of visual sequence prompts (a series of client-viewpoint photographs of correct steps) was tried for six limb donning attempts. He was unable to benefit from these. We suspect that this was due to the attentional demands of shifting attention from current task to the visual cues.

*Task:* A standardized sequence was developed for prosthetic limb donning, following interview with physiotherapists and prosthetists. This was then instantiated as a
computer-stored sequence GUIDE© and delivered to the patient via wireless headphones.

*Design:* A single participant baseline-intervention experiment was employed. Baseline data were gathered by video recording on-task behaviour. These videos were scored, using a customised Sequence Performance Scale, for a. percent correct b. errors of omission, c. deviations from sequence, d. repetitions and e. time to fit limb to give data points for statistical analysis (Semkovska et al., 2003). Inter-rater reliability for this measure ranged from 0.73 to 0.77 across the indices of the measure. The intervention trials were randomly allocated within the sequence of trials.

*Results:* The participant’s performance scores by trial are shown in Figures 1 and 2. On the x axis one can see the trial number, separated into baseline on the left and intervention on the right. BB was on average 81% correct in the baseline and 94% correct on the intervention trials. An exact probability test for randomized trials (Todman & Dugard, 2001) found that GUIDE© use was associated with statistically significant reductions in repetitions of previous steps (p<0.05) and omissions of sequence steps (p<0.05). Reduction in deviations from the ideal sequence approached significance (p=0.061). Time to don the prostheses was reduced from a mean of 12 minutes in baseline to 9.25 minutes on GUIDE© trials. An Efficiency Index (Total Correct / Total Time) also changed significantly between conditions (p<0.05).

[Insert Figures 1 about here]

[Insert Figure 2 about here]
Discussion

Both of our preliminary studies show benefits of using a voice mediated prompting system over sequence performance without. This was in the context of imposed distractors in Study 1 and in a participant with cognitive impairment in Study 2. Taken together, these studies show the feasibility of using an auditory-verbal interface in ATC design for contexts of high distraction and cognitive impairment.

The GUIDE© demonstrates potential as a rehabilitation tool. Cognitive problems compromise rehabilitation outcome in conditions such as cerebrovascular accident (Paolucci et al., 1996) and post-amputation (O’Neill, 2008). For many of these conditions rehabilitation is labour intensive due to the need for repeated instruction or supervision to prevent errors. Computer assisted guidance of repeated instruction could thus allow patients to engage autonomously in rehabilitation relevant tasks.

The success of the GUIDE©, we suggest, stems from the verbal interface. While the mode of verbal interaction is novel for ATCs, it is familiar for users. BB adapted to the use of the GUIDE© in the first session. While many other ATCs have reported difficulties in training users and long learning curves (LoPresti, Mihailidis & Kirsch, 2004), we suspect this has been due to the lack of familiarity with the interface. A second advantage with the verbal interface, which we argued in the introduction, is that it provides a relatively direct route to augmenting executive function because there is evidence to suggest that executive function is in fact heavily verbally
mediated. The success of the GUIDE© is consistent with such an association.

Moreover, during the trials when BB was not using the GUIDE© the researchers noted an increased tendency for BB to talk himself through the task. This talk was loud enough to be heard by the researchers. Such utterances are in line with research which has shown that children facing a complex task tend to talk themselves through the task (Berk & Garvin, 1984; Vygotsky & Luria, 1994; Winsler & Naglieri, 2003), thus again supporting the link between executive function and verbal mediation which the GUIDE© is based upon.

All users in the above studies showed good tolerance of the system. In Study 1 more positive comments were made with regard the GUIDE© than the written instructions. It is hypothesised that a verbal guidance system is tolerated because the delivery of prompts is similar to the experience of private speech, which underpins normal problem solving (Phillips, Wynn, Gilhooly, Della Sala & Logie, 1999). BB commented after some trials using the GUIDE© ‘how did I do?’ or ‘I did well there’. This contrasted with his previous thanking of the instructor or making disparaging comments about them in the event of poor performance. It thus appears likely that persons using the system may be able to attribute success to the actions of the self rather than use of the system. This is promisingly indicates that users may experience personal satisfaction attaining a goal using the system, despite being scaffolded to do so.

The acceptability of having a sequence step cued by a verbal stimulus appears good. What is less clear is the acceptability of making verbal responses to control cue deployment. Each of the participants performed the focal task alone apart from the
presence of the experimenter and it remains a possibility that users may feel self-conscious if using the system in earshot of others. Future research is needed to examine the extent to which using the device in public settings is perceived as stigmatising.

The emulation of carers’ verbal guidance of behaviour in the field of assistive technology for cognition is rare. Orpwood, Adlam, Gibbs and Hagan (2001) used recorded carers’ voices to remind persons with dementia of omissions triggered by sensors. For example a bath sensor triggered the recording: “Don’t forget you’ve left the bath running mum”. Literature search did not elicit any research assessment of the effectiveness of such devices although the initial case studies are promising. To our knowledge our system is the first to accept users’ verbal responses to control output. We would argue that some control of the output is an important feature. Being unable to control the voice directing behaviour may be noxious and disempowering. Perception of an assistive technology is an important determinant of use (e.g. Scherer, 2002) and we argue that the acceptability of devices (e.g. monitoring systems, prompters etc.) should be to the fore of any assessment of their efficacy.

The studies reported here demonstrate the orthotic function of the system. Performance-with was only compared to performance-without and there was no examination of learning effects. However, for individuals with better memory function the system might also restore performance of complex sequences. Errorless learning, wherein performance errors are minimised for the memory impaired learner, has evidence to suggest efficacy in teaching ordered semantic information (Kessels & De Haan, 2003) and behavioural sequences (Maxwell, Masters, Kerr & Weedon,
The system can guide error free performance and might thus be used to support errorless learning. Directions can be gradually replaced by question prompts with these faded out in turn, incrementally increasing the amount to be recalled.

The range of potential applications is difficult to estimate but appears wide. All complex behavioural sequences which can be verbally described are open to the application. For future research we suggest identifying and then augmenting those behaviours which formal and informal carers are currently providing verbal scaffolding for. In depth analysis of carer providers as ‘assistants for cognition’ might provide further clues as to the most appropriate modes of questioning and prompting. Also studies of the self-talk that normally accompanies the performance of a complex behavioural sequence are needed. The motivation for this research is to construct the simplest and most intuitive interface for ATCs. To this end, the future development of ATC interfaces, we suggest, should focus upon the auditory-verbal interface to simulate both carer instruction and the internal verbal control of sequence performance.

References


Tables and Figures

Table 1 – Mean and Standard Deviation Scores for Key Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>The Guide (n=13)</th>
<th>Written Instructions (n=13)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Percentage of Correct Steps</td>
<td>91.03</td>
<td>12.16</td>
</tr>
<tr>
<td>Percentage of Incorrect Steps</td>
<td>8.97</td>
<td>12.16</td>
</tr>
<tr>
<td>Number of Hesitations</td>
<td>1.35</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Mean time for each Step

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.15</td>
<td>0.29</td>
<td>1.29</td>
</tr>
<tr>
<td>Positive Comments</td>
<td>2.50</td>
<td>1.31</td>
</tr>
<tr>
<td>1.50</td>
<td>1.31</td>
<td>1.50</td>
</tr>
<tr>
<td>Negative Comments</td>
<td>1.50</td>
<td>1.51</td>
</tr>
</tbody>
</table>

Figure 1: Sequence performance repetitions by condition

![Bar chart showing sequence performance repetitions by condition](image)

Figure 2: Sequence performance omissions by condition

![Line graph showing sequence performance omissions by condition](image)