Computer delivery of gesture therapy for people with severe aphasia

Jane Marshall[1]

Abi Roper[1]

Julia Galliers[2]

Stephanie Wilson[2]

Naomi Cocks[1]

Sam Muscroft[2]

Tim Pring[1]

[1] Division of Language and Communication Science, City University London

[2] Centre for Human Computer Interaction Design, City University, London

Corresponding Author: Jane Marshall, Division of Language and Communication Science, City University London, EC1V OHB. **J.Marshall@city.ac.uk**

Abstract

Background: Using gesture as a compensatory communication strategy may be challenging for people with severe aphasia. Therapy can improve skills with gesture, at least in elicitation tasks, but gains are often modest (e.g. Daumuller and Goldenberg, 2010; Marshall, Best, Cocks, Cruice, Pring, Bulcock, Creek, Eales, Lockhart Mummery, Matthews & Caute, 2012). Raising the treatment dose with technology might improve outcomes.

Aims: This feasibility study developed a computer gesture therapy tool (GeST), and piloted it with nine people who have severe aphasia. It aimed to determine whether practice with GeST would improve gesture production and/or spoken naming. It also explored whether GeST encouraged independent practice and was easy to use.

Methods and Procedures: Pilot participants had 6 weeks practice with GeST, flanked by pre and post therapy tests of gesture and word production. Usability was explored through interviews and structured observations, and the amount of time spent in the programme was monitored.

Results: Scores on the gesture test were evaluated by 36 independent raters. Recognition scores for gestures practised with the tool improved significantly after therapy and the gain was maintained. However, gains were small and only occurred on items that were practised with regular therapist support. There was no generalisation to unpractised gestures and no effect on spoken naming. Usability results were positive. Participants undertook an average of 64.4 practice sessions with GeST, and the average session length was just under 14 minutes.

Conclusions: GeST proved to be easy and enjoyable to use and had some effect on participants' gesturing skills. Increasing the magnitude of gains would be desirable. The effect on everyday communication needs to be explored.

Introduction

In the face of severe aphasia, gesture is an obvious candidate for communication. It requires no additional resources, such as pen and paper, and is a typical and universal human behaviour (Kita, 2009). Definitions of gesture typically draw a distinction between those gestures that accompany speech and those that convey meaning in isolation (e.g. McNeil, 2005). The latter include pantomime gestures, which is the type of gesture focussed in this study.

The literature contains several accounts of people with aphasia who revealed remarkable skills with gesture, in both interactive communication (e.g. Goodwin, 2000; Parr, 2007; Wilkinson, Beeke & Maxim, 2010) and formal testing (Kemmerer, Chandrasekaran & Tranel, 2007; Marshall, Atkinson, Smulovitch, Thacker & Woll, 2004). The case of Charles (Marshall et al, 2004) was particularly striking. He was a deaf user of British Sign Language who had aphasia following a stroke. Sign finding was very impaired, with hesitations, omissions and phonological errors. Charles was observed to use gesture in conversation to compensate for his sign finding difficulties. Formal tasks also showed that he could gesture the use of objects much more readily than sign their names. Furthermore, this contrast was evident even when signs were highly iconic and superficially very similar to their corresponding gestures. Thus a very sharp dissociation was demonstrated between impaired access to word forms, and intact access to gestures, despite the fact that, in this case, language and gesture shared the same modality.

However, such dissociations are by no means universal. A number of studies have found that gesture and language impairments often go hand in hand, with parallels in severity and type (e.g. Cicone, Wapner, Foldi, Zurif & Gardner, 1979; Duffy, Duffy & Mercaitas, 1984; GlosserWeiner & Kaplan, 1986). For example, Cicone and colleagues (1979) found that people with non fluent aphasia produced infrequent but clear gestures alongside speech; while people with fluent aphasia produced copious but largely uninterpretable gestures.

The reason for gesture impairments in aphasia is disputed. Parallels with language symptoms gave rise to the proposition of a central asymbolia, which equally impairs access to words and representational gestures (Duffy & Duffy, 1981; Goldenberg, Hartmann & Schlott, 2003). Others have appealed to more specific aspects of cognition, such as non verbal semantics (Hadar, Burnstein, Krauss & Soroker, 1998; Hogrefe, Zeigler, Weidinger & Goldenberg, 2012). Purdy and Koch (2006) argue that impaired executive functioning inhibits the problem solving skills needed for gesture use. They suggest that shifting from speech to a non verbal modality imposes heavy executive demands that cannot be met by some people with aphasia.

A further explanation for gesture impairments is the presence of limb apraxia, which inhibits both the planning and execution of hand movements. In line with this, Hogrefe et al (2012) found that apraxia scores correlated with the diversity and intelligibility of pantomime gestures in 20 people with severe aphasia. However, there is also evidence that spontaneous gesturing can proceed despite the presence of apraxia (e.g. Rose and Douglas, 2003).

Despite the contradictions in the literature it seems that the ability to use gesture in aphasia cannot be assumed. Rather, for many individuals this will need therapeutic support. Furthermore, treatments should pay attention to the various proposed reasons for gesture impairments, e.g. by underscoring the meaning of gestures and/or giving intensive opportunities to practise hand movements. Targeting non verbal problem solving skills, such as modality shift, may also be required for those who exhibit executive difficulties.

Early attempts to exploit gesture in therapy claimed some remarkable successes (see Rose, 2006 for review). For example, Skelly, Schinsky, Smith & Fust (1974) taught Amer-Ind signs to 6 people with oral apraxia. All participants mastered 50 signs within 2 months. However, sign acquisition was not formally tested but was rather observed in a range of topic based therapy tasks. Rose (2006) argues that such a lack of experimental design bedevilled early studies of gesture therapy.

Daumuller and Goldenberg (2010) addressed this problem, by comparing outcomes across treated and control participants. Treated participants were taught 24 communicative gestures over 3 treatment phases, while controls underwent repeated testing with no intervening therapy. The test procedure required participants to produce a gesture in response to a stimulus object or verbal request. Results showed that gesture production improved in the treated group, but not in the controls. Gains were mainly confined to treated gestures, although there was some weak generalisation to untreated ones. Despite these significant results, the rate of learning was much less impressive than apparently achieved by Skelly et al (1974), with 3 hours of therapy needed for each newly acquired gesture. The authors also acknowledge that their design did not explore the communicative use of gestures.

Daumuller and Goldenberg (2010) targeted gesture purely as a substitute for speech. Others have explored its potential as a speech facilitator (e.g. Boo & Rose, 2011; Marangolo, Bonifazi, Tomaiuolo, Craighero, Coccia, Aleoe, Provinciali & Cantagallo, 2010; Raymer, Singletary, Rodriguez, Ciampitti, Heilman & Rothi, 2006; Rodriguez, Raymer & Rothi, 2006; Rose & Douglas, 2008). These studies suggest that naming therapies incorporating a gestural cue can significantly enhance both noun and verb production. However, as treatments combined verbal and gestural elements, it is difficult to determine the independent contribution of gesture.

A recent study conducted at City University London aimed to explore a number of the unresolved questions surrounding gesture therapy in aphasia (Marshall et al, 2012; Caute, Pring, Cocks, Cruice, Best & Marshall, in press). In the first phase of the study, fourteen people with severe aphasia received therapy aiming to train 20 gestures and 20 (different) words. Gesture and naming therapy were conducted separately, but in parallel, using similar elicitation and cueing techniques. The primary outcome measures were tests of naming and gesture production, administered twice before and twice after therapy. Each test contained sixty items, of which 20 had received naming therapy, 20 had received gesture therapy and 20 were untreated. Scoring of gestures evaluated the degree to which they could be recognised, and was conducted by assessors who were blind to the treatment targets and to the time of assessment.

Results showed that participants made significant gains both in gesture and word production as a result of therapy. Treatment effects were item specific; i.e. gains in gesture occurred largely on items that received gesture therapy and gains in naming occurred on items that received naming therapy. There was no evidence of cross modality cueing. In other words, items that were treated for gesture were named no more successfully after therapy than before; and items that were treated for naming showed no gains in gesture. Encouragingly, there was some evidence that the gestural and naming gains made by participants benefited their interactive communication, as there were improvements on tasks in which they had to convey messages and narratives to their partners.

Like Daumuller and Goldenberg (2010), Marshall et al were interested in the rate and extent of learning. Interestingly, they found that naming gains significantly outstripped gesture gains. So, on average, their participants acquired 8 new words, compared to just under 2 new gestures following 7.5 hours of therapy in each modality. The authors propose various reasons for the modest gestural gains. They make the point that all the participants had limb apraxia, which may have limited progress. They also argue that gesture and naming pose unequal learning demands, in that gesture therapy has to teach novel forms, while naming therapy has to restore access to forms that were familiar prior to the stroke.

The small number of participants in this study made it difficult to explore baseline predictors of gain. For example, scores on an apraxia screening test and a semantic memory test showed no relationship with therapy outcomes. However, subsequent analyses have brought one significant correlation to light. Unforeseen circumstances, such as ill-health and family commitments, meant that the time taken to complete the therapy varied across participants, and this was found to correlate inversely with gesture (but not naming) therapy gains. Caution is required when interpreting this finding, not least because those who postponed sessions may have been the poorest candidates for treatment. Nevertheless it may suggest that intensity of practice is particularly crucial for the learning of gestures, e.g. because of their specific learning demands, or because apraxic barriers have to be overcome.

One way of achieving a high intensity therapy dose is by engaging technology. Several self administered, computer based treatment packages have now been developed for people with aphasia, with positive outcomes for sentence production (e.g. Linebarger, McCall, Virata & Berndt, 2007), auditory comprehension (Archibald, Orange & Jamieson, 2009), speech apraxia (Whiteside, Inglis, Dyson, Roper, Harbottle, Ryder, Cowell & Varley, 2012) and naming (Laganaro, Di Pietro & Schnider, 2006; Ramsberger & Marie, 2007; Palmer, Enderby, Cooper, Latimer, Julious, Paterson, Dimairo, Dixon, Mortley, Hilton, Delaney & Hughes, 2012). These studies also confirm that electronic delivery can successfully raise the treatment dose. For example, Palmer et al (2012) found that just under 9 hours of therapist/volunteer support enabled an average of 25 hours independent practice, and participants in Whiteside et al (2012) carried out an average of 43 self directed computer based treatment sessions over 6 weeks. However, there is also a high degree of individual variation in the amount of practice undertaken. For example, the nine participants in Archibald et al (2009) had computer usage times ranging from 5.51 to 39.47 hours; and Linbarger et al (2007) report practice periods ranging from 11 to nearly 70 hours.

To date, computer therapy tools for people with aphasia have focussed purely on spoken or written language. Compensatory modalities, such as gesture, have not been targeted. Yet recent technological advances in motion and gesture recognition make this possible. A recent research project at City University London aimed to build a computer therapy tool for gesture using such ‘mainstream’ technologies. The tool, named ‘GeST’, was designed for people with severe aphasia, for whom the compensatory use of gesture is likely to be recommended. It was built through a process of participatory design, in which representatives of the end user group (in our case, 5 people with aphasia) inform all decisions. It was then piloted with a small group of (different) participants to explore its usability and therapeutic benefits.

This paper briefly reports the design process of GeST (see also Galliers, Wilson, Muscroft, Marshall, Roper, Cocks & Pring, 2011; Galliers, Wilson, Roper, Cocks, Marshall, Muscroft & Pring, 2012) and describes the resulting tool. It then outlines the methods and results of the pilot study. While very preliminary, the results of this pilot suggest that practice with GeST enhanced gesture production in 9 participants with severe aphasia. However, gains were modest, item specific and seemed dependent on regular therapist support. Next steps in the evaluation of GeST are proposed.

Design of the Tool

In building GeST we aimed to create a therapy tool that could be used autonomously by people with severe aphasia, and which offered opportunities for graded learning of gestures. We also wanted to motivate intensive practice, e.g. through positive feedback and an attractive and enjoyable format.

To meet these challenges, we used participatory design. Five people with aphasia were employed as consultants to trial and give their feedback on different technological options. Their views were elicited through a series of design workshops, each of which was attended by 2 or 3 consultants, a Human Computer Interaction (HCI) researcher, a technology researcher, and a Speech and Language Therapist (SLT) researcher. In each session, a specific technology or mode of presentation was introduced and tested by the consultants. They were observed using the technology and their views were subsequently canvassed by an interview at the computer, e.g. so that they could directly indicate successful features. Interviews were conducted by the SLT researcher and employed 'aphasia friendly' techniques, such as yes/no questions and simplified rating scales.

Every design decision about GeST was informed by this process. For example, early sessions were used to select between vision-based and motion-sensing gesture recognition technologies. Later sessions explored presentation and navigation options.

These, for example, showed that a 3D gaming environment was positively received by consultants, and helped to refine the simplified keyboard used in GeST.

The final prototype of GeST is shown in Figure 1 a & b. It is programmed with 30 gestures, each presented in three therapeutic levels. In the first level, participants see a therapist modelling a gesture and are invited to repeat that gesture. So they see a picture of the target and hear: ‘Here is a gesture for (name of target). Now it’s your turn’. In the next level they see a computer character (Gerry) walking through a 3D virtual environment. For example, one such environment is a park. Events happen in the environment, which stimulate the need for an object. The gesture for that object is then introduced by the therapist and participants are invited to repeat the gesture. Once their gesture is recognised by the computer, Gerry is shown to acquire the object. To give an example, in the park it starts to rain. The gesture for umbrella is modelled, and after this is produced by the participant Gerry is shown holding an umbrella. The third level demonstrates the use of the target gestures in real world scenarios. Video clips taken in everyday settings show people using gesture to communicate with one another. After the scene, participants are invited to repeat the gesture, before moving on to the next scene. For example, one scenario shows a man about to leave a house, dressed in a coat, hat and scarf. He pauses by the door and gestures glove. His partner says ‘gloves?’ and hands them to him. Participants are then invited to repeat the gesture for glove. The actors in the scenarios included two of the project consultants. Thus people with and without aphasia are shown using gesture.

Insert Figure 1 about here

In all three levels participants are cued to produce gestures by a 3-2-1 countdown on the screen and by a bell. When the bell sounds, the gesture recognition becomes operational. The web cam mounted on the laptop detects the participant's handshape and images this on the screen. When the handshape matches the gesture target recognition is acknowledged by an auditory round of applause. In the second level, participants are also rewarded by a pile of money imaged at the top of the screen, which increases with each gesture recognised. The vision-based technology is configured to detect the colour yellow. Participants therefore wear a yellow cotton glove while using GeST. Interestingly, none experienced any problems in getting this on and off, despite their hemiplegia[1].

Using GeST involves just six buttons (see figure 1). The prominent green 'on' (power) button brings participants directly into the home screen of the programme. Participants then select their preferred level of practice from three images, by clicking the large yellow 'ok' button. Navigation between these levels is via the left and right arrows. Movement between gestures within each level is achieved by pressing 'ok'. Participants can return to the home page by pressing the 'menu' button. The computer can be shut down at any point by pressing the red 'off' button (the modified 'enter' key).

Finally, GeST is configured only to provide positive feedback, when gestures are recognised. Inaccurate gestures, or gestures that are not recognised by the technology, elicit no response from the computer. Users are therefore encouraged to repeat their gestures until recognition occurs. They also have the option of moving the programme on to a new target by pressing the forward arrow.

The Pilot Study

The pilot study received clearance from the Ethics Committee of the School of Health Sciences, City University London. All participants gave informed consent.

The study aimed to address the following questions:

- Will practice with GeST improve participants' production of gestures &/or the corresponding spoken words?
- Will improvements be specific to the 30 items that feature in the programme?
- Will gains occur when GeST is used without ongoing therapist support?
- Will gains be maintained after GeST is withdrawn?
- What are participants' views about GeST?
- What are their partners' views about GeST? (where relevant)
- Is GeST easy to use in the home setting?

We were also interested in the dosage delivered by GeST, and whether this affected therapy gains. Quantitative usage data were collected automatically by the software; i.e. the number of practice sessions conducted each day, the length of sessions and the amount of usage at each level of the tool. The number of gestures recognised by the tool was also recorded, although only at level 2.

Ten people with severe aphasia were recruited to the pilot study. One failed to complete owing to ill health, so data on nine are reported. All were fluent pre-stroke users of English, established via self and relative report, although three were bilingual. They had severe aphasia, with a score of below 20% on the spoken naming subtest of the Comprehensive Aphasia Test (Swinburn et al, 2004). An object to picture matching screening test established that they could recognise pictures. They were also screened for limb apraxia (Dabul, 2000), with three showing evidence of impairment. None had a diagnosis of dementia or any other progressive neurological impairment. Most participants lived with a spouse or family member(s), but one lived alone. An informal case history established their previous use of technology. Their details are reported in Table 1.

Insert Table 1 here

Participants received 6 weeks practice with GeST, delivered in two phases. In the first phase they were supported by weekly visits from the SLT researcher (2nd author). In the second phase they used GeST without therapist support. Each phase consisted of three one week blocks, and during each block five gestures were practised on GeST, potentially at each level. Therefore a total of 15 gestures were practised in phase 1 with therapist support, and 15 different gestures were practised in phase 2 without therapist support.

At the start of therapy, GeST was set up in participants' homes. Positioning aimed to achieve best lighting conditions, given that gesture recognition was vision-based, and paid careful attention to safety and convenience. Each participant received a demonstration session, which familiarised them with the key features of GeST. So they practised using the on/off buttons and were shown how to use the simplified key board to navigate between gestures and levels. Participants were then left to practise the first block of gestures independently. In phase 1 the SLT researcher visited at the end of each week to review progress. She observed the participant's use of the tool from start to finish and modelled areas of computer use which required additional support. For example, the use of the "home" and "OK" button to switch between levels might be practiced and reinforced. Gesture accuracy was also targeted. If a participant's gesture was unclear, it would be re-modelled with relevant feedback and then practiced by the participant until GeST was able to identify it. Gestures which were recognised adequately by GeST, but which the researcher deemed could be further refined to improve intelligibility for a human observer, were also demonstrated and reinforced.

There was one additional component. The 30 gestures practised across both phases of GeST were initially familiarised. So the therapist showed the participant a picture of the target item and demonstrated the gesture that would feature in GeST. This was done to encourage the use of standard, rather than idiosyncratic gestures. During phase 1, fifteen gestures that would *not* feature in GeST were additionally familiarised. So, at the start of each week ten gestures were familiarised, of which five featured in the programme. This element of the design enabled us to investigate whether familiarisation alone brought about learning effects, or whether learning depended on practice with GeST.

The study employed a repeated measures design, with testing of treated and control items. Tests were administered before therapy (time 1) after phase 1 (time 2), after phase 2 (time 3) and after a three week maintenance period in which participants had no further practice with GeST (time 4). Tests were administered live by the second author and required participants to name and, on a separate occasion, gesture 60 object pictures. In the naming condition they were shown each picture in turn and asked: 'Here is a picture. Try to say the name of the picture. Use one word.' In the gesture condition they were shown the same picture and asked: 'Here is a picture. Try to think of a gesture for the picture. Show me the gesture.' Testing was conducted over two sessions, with half the items named and half gestured in each session. Test order (naming vs gesturing) was counter balanced over sessions. All tests were video recorded.

The sixty test items comprised four groups. Fifteen items were practiced in phase 1 of GeST (supported GeST items), fifteen were practised in phase 2 (independent GeST items), fifteen were familiarised in phase 1, but not practised with

GeST (familiarised items) and 15 were neither familiarised nor practised (control items).

Gesture responses were evaluated by scorers who were not members of the research team (most were students of speech and language therapy). Each participant's gesture responses were edited onto 4 scoring tapes. Tapes contained all 60 items but drawn from the four testing occasions. So, fifteen gestures were produced at time 1, fifteen at time 2, fifteen at time 3 and fifteen at time 4. Gestures were presented in random order to ensure that scorers were blind to the time of testing. Each scorer evaluated only one tape, so saw each gesture target only once. This procedure ensured that differences between scorers could not affect the evaluation of change over time, since scorers evaluated an equal number of gestures across all the time points.

Scorers saw a video clip of each gesture and were asked to write down their understanding of the gesture target. To gain a point for recognition, scorers had to name the target item or a synonym for that item as listed in the 2007 version of Microsoft Word. After scorers had completed the recognition task the targets were revealed. They were then asked to rate the accuracy of the participant's gestures on a 1 - 5 scale, with 5 being 'Very accurate' and 1 being 'Not at all accurate'. This procedure yielded a recognition and a rating score for each gesture at each time point. The dual scoring aimed to be maximally sensitive to change. It was hypothesised that even well executed gestures might be difficult to recognise without any context. For example, the target gesture for glove was a hand held up with extended fingers and thumb. While this might be understood as glove, other interpretations are possible (such as 'stop' or 'good bye'). Therefore, improvements in the production of this gesture might be missed in the recognition score but detected in the rating score.

Reliability of the scoring system was checked by double scoring a subset of the data. Three participants were double scored for recognisability and two for rating. Percentage agreement for recognition scores was good (87.2%). Spearman correlation coefficients were used to explore the agreement between rating scores. These yielded values of .715, $p<.001$ and .475, $p<.001$ across the pairs of scorers.

Naming responses were transcribed and scored by the second author. To be correct participants had to name the target or a synonym. Phonological errors that were recognisable attempts at the target were scored as incorrect.

Participants' views on GeST were explored through interviews conducted by the HCI researcher (the 3rd author) at the final time point, using a standard and identical set of questions. Participants were asked about enjoyment, likes and dislikes and navigation. Responses were facilitated by using a yes/no question format and a simple opinion rating scale. Where possible, family members or friends were similarly interviewed, again using standard questions. Topics included whether or not the participant had used GeST independently, any challenges experienced, observed benefits e.g. on everyday communication, and negative aspects, such as carer burden.

Mastery of the computer interaction features was assessed through structured observations conduced by the HCI researcher. These were conducted soon after GeST was installed (Observation A), after phase 1 (Observation B) and after phase 2 (Observation C). Each participant was observed carrying out a GeST session and 8 behaviours were assessed, such as using the on/off buttons, navigating between levels with the arrow keys, and using OK to progress between gestures. Participants scored 1 if the behaviour was used consistently; .5 if the behaviour was used at least once; and 0 if it was not demonstrated.

Results of the Pilot Study

Gesture Recognition Scores

Gesture recognition scores across the 4 time points (see Figure 2 and Table 2) were analysed with a two factor, within subject ANOVA. Factors were time (4 levels: time 1, time 2, time 3 and time 4) and group (4 levels: supported GeST, independent GeST, familiarised and control).

Insert Figure 2 and Table 2 about here

Neither of the main effects was significant, but there was a significant interaction between time and group F (9, 72) = 2.12, $p<0.05$; $\eta^2 = .210$.

To explore the origin of this interaction, separate one factor ANOVAs were conducted on each group of items. Only items that received supported GeST showed a significant effect of time, F (3, 24) = 3.27, $p<0.05$, $\eta^2 = .29$. Planned comparisons for this group, using related t tests, showed that scores at time 2 differed significantly from scores at time 1 (t (8) = 4.15, $p<0.005$). There was no difference between time 2 and time 3, or between time 3 and time 4. A comparison between times 1 and 4 showed that scores at the final assessment were still significantly higher than at the outset (t (8) = 2.31, $p<0.05$).

Thus items that were practised in GeST with the therapist's support improved, and this improvement was maintained. None of the other groups showed any significant improvements. Gains for the supported GeST items were spread fairly evenly across the group (see Appendix for the list of items treated in GeST with the per item gain scores).

Gesture Rating Scores

Gesture rating scores (see Figure 3 and Table 2) were analysed with a two factor within subject ANOVA. Factors were time (4 levels: time 1, time 2, time 3 and time 4) and group (4 levels: supported GeST, independent GeST, familiarised and control). Neither of the main effects, nor the interaction was significant. Figure 3 shows that rating scores were remarkably consistent across the groups and across time.

Insert Figure 3 about here

Naming Scores

Naming scores (see Figure 4) were analysed with a two factor within subject ANOVA. Factors were time (4 levels: time 1, time 2, time 3 and time 4) and group (4 levels: supported GeST, independent GeST, familiarised and control). Neither of the main effects, nor the interaction was significant. Figure 4 shows that naming scores were at or close to floor for all groups at all time points.

Insert Figure 4 about here

Interview findings: Participants

Eight of the nine participants rated their enjoyment of GeST highly or very highly. The remaining participant pointed to her mouth rather than the rating icons, possibly indicating that her priority for therapy was speech. There were varying views about the preferred levels of practice. Three participants expressed a preference for level 1, and two for level 2. Two participants liked level 2 and 3 equally; and the remaining two rated all levels equally.

The feedback provided by GeST was well received. Seven participants indicated that they enjoyed the clapping and four expressed positive views about the monetary rewards accrued in level 2 (three participants even kept a record of their scores). Five participants, however, were indifferent to this feature. The majority of participants (six) indicated that they felt in control of GeST, although two gave low ratings for their mastery over the programme.

Interview findings: Partners

Eight partners were available for interview. Of these, seven indicated that participants used GeST entirely independently. The remaining partner said that the participant was 80-90% independent, but needed help initially. Most partners (7) noted that participants used GeST without prompting, e.g. taking themselves to the computer. All reported that GeST was enjoyable and simple to use. e.g.:

> *'I was a technophobe, and when they said 'computer' I thought it was going to cause problems. I thought I wouldn't understand and he wouldn't understand it. But it's so easy.'*

Few challenges were identified. Two partners noted that the gesture recognition was variable, and two recalled instances when the computer failed to shut down. One partner indicated that navigation between levels was challenging.

When asked about benefits, several partners commented that use of GeST and/or the experience of mastery over the programme had very positive impacts on the participant's mood, e.g.:

> *'He is more confident. Eager. Not lethargic. Wanting to show anyone who comes'*

Four partners felt that participants were using more words or gestures. Although two expressed reservations about whether gestures practised on the computer would generalise to everyday life, e.g.:

> *'It would be helpful if there could be a more focussed way of helping the person doing it to integrate the gestures into ordinary life ….While she works on it here [points to computer], it doesn't necessarily translate.'*

None of the partners found GeST intrusive in the home or burdensome for them.

Observation findings

Observation data are reported in Figure 5. These show that many of the navigational features of GeST were at least partly accessible to participants even at the first observation (A), and that mastery increased over the practice periods (B & C). Navigation between levels was clearly more demanding than navigating within levels, in line with one of the partner's comments. The use of arrow keys to re-start a gesture or move on was the least used aspect of the programme. However, this was an optional control of pace. Progression to the next gesture occurred automatically if participants did not use this feature.

Insert Figure 5 about here

Usage Results

Usage data were collected automatically by GeST, enabling us to determine the amount of practice undertaken by each participant, whether this changed over the two phases and the preferred level of practice.

On average, participants used GeST for 13.9 hours over the two phases (range 7.56 – 39.26 hours). The mean number of sessions across the group was 64.4 (range 20 – 95) and the average session length was just under 14 minutes (range 7 – 25 minutes). Thus participants tended to use GeST in frequent, relatively short bursts. The ranges suggest considerable individual variation. However, this was largely due to one outlying participant (see Figure 6).

There was a substantial, although not significant difference between usage in phase 1 and phase 2 (mean usage in phase 1 = 8.03 hours; mean usage in phase 2 = 5.86 hours; $t = 1.85$, $p > 0.05$). Figure 5 shows that most participants practised less in phase 2, with the exception of participants 3 and 9.

Insert Figure 6 about here

Figure 7 shows usage across both phases broken down by level. Varying patterns are demonstrated, with some participants showing a preference for one level (1, 3 & 6) and others showing a more even profile (2, 7 & 8).

Insert Figure 7 about here

Finally we explored the relationship between usage and gain. The number of gestures acquired by each participant across both practice phases with GeST was calculated, and correlated with their overall usage time, using Pearson Correlation Coefficient. Each person's level 2 recognition score (the number of gestures recognised by the GeST software) was also correlated with the gain scores. Both results were insignificant ($r = .29$; $r = .23$).

Discussion

This project aimed to develop and pilot a new computer gesture therapy tool for people with aphasia. Development was through a process of participatory design, ensuring that all technological decisions were informed by representatives of the end user group. The resulting tool enabled users to practice 30 gestures, with confirmatory feedback following correct gesture production. GeST involved just six operational buttons, so allowed autonomy of use, even for participants with severe strokes. The three therapy levels offered hierarchical practice, and exemplified gestures in different contexts and environments. So level one modelled the target gestures, level two presented the gestures within a 3D 'gaming' environment and level three demonstrated their use in real world scenarios.

The pilot study aimed to determine the therapeutic benefits of 6 weeks practice with GeST, participants' views about the tool and its ease of use. We were also interested in the amount of practice undertaken by participants, i.e. the self administered therapy dose.

Taking the experimental results first, the gesture recognition scores suggested that using GeST with weekly support from the SLT researcher improved the production of gestures that featured in the tool. These gestures were recognised by scorers after the practice period more successfully than before, and this gain was maintained at the final assessment. Using GeST entirely independently, however, did not result in significant gains. Improvements were also item specific, with no generalisation to gestures that were not included in GeST. Indeed recognition scores on the untreated control items marginally reduced over time. There were also no benefits from using GeST for spoken naming. This is in line with previous research evidence. Benefits for naming have been observed when gestural cues are integrated into treatments for word production (e.g. Marongolo et al, 2010). However, targeting gesture in isolation often fails to stimulate speech, particularly when the aphasia is severe (e.g. Marshall et al, 2012). It also suggests that merely hearing spoken words, as occurred in GeST, is not sufficient to stimulate naming.

Several limitations in these results need to be acknowledged. Firstly, the improvement achieved in the supported GeST items was very modest; i.e. the average gain in recognition scores between time 1 and time 2 was just under two items, and this fell back to 1.5 items at the follow up (time 4) assessment. Secondly, although the supported GeST group was the only one to change significantly over time, differences were again marginal. Figure 2 shows that items that were merely familiarised in the first phase also improved, albeit not significantly. There is also the puzzling finding that while recognition scores changed, rating scores did not. Rating scores were collected in addition to recognition scores in an attempt to increase sensitivity. It seemed, however, that ratings for accuracy, in fact, failed to detect subtle changes that made gestures more recognisable. Finally, our data do not throw any light on whether GeST improved the use of gesture in everyday communication. Indeed, two of the partners expressed reservations about the generalisation of skills.

Turning to our interviews, these revealed very positive views about GeST, both from participants and partners. The tool was found to be accessible and enjoyable to use, and highly motivating. For example, participants liked the feedback provided by GeST, and partners noted that practice was typically self initiated. Reassuringly, GeST did not add to carer burden, which is important for a home based self administered treatment, and few technological glitches were identified.

The interview data on usability were corroborated by the results of our structured observations. All participants displayed consistent or occasional use of most navigational aspects of GeST. Navigation between levels using arrow keys was demonstrated less consistently than navigation within levels using the OK key, which may point to the need for more initial training on this aspect.

Usage data revealed a pattern of ‘little and often’. The mean number of sessions was 64.4, indicating that participants typically practised every day, but with an average session duration of just under 14 minutes. One ‘uber user’ bucked this trend. He practised for a total of 39.28 hours, with sessions often lasting for nearly 30 minutes. It seems, therefore, that participants made personal choices about their preferred practice regimes. We examined whether patterns of usage related to benefit, by correlating individual’s usage time with their gain score. Results were not significant, which is perhaps not surprising given the very small number of participants.

There was a marginal, although non significant, decline in use during phase 2. This finding invites some re-appraisal of the experimental findings. Items practised in phase 1 (supported GeST) were the only ones to show a significant gain. These items may have benefited from the regular support of the SLT researcher or a higher practice dose. It is important to stress that the usage data for phase 1 included the three sessions that were supported by the researcher, each lasting one hour. Given that the mean difference in usage between phase 1 and phase 2 was 2.17 hours, these sessions more than account for the difference. Importantly, this indicates that there was no decline in the amount of self directed practice carried out in phase 2.

Our usage data also enabled us to examine the amount of practice conducted at each level of the programme. Here again there was evidence of personal choice. For example, some participants practised mainly at level 1, while others preferred level 2 or practised equally at all levels. In some cases, these patterns reflected the preferences expressed in the interviews; e.g. participant 6, who mainly practised level 1, indicated that this was her favourite level. However, it is also worth noting that navigation between levels was the most challenging aspect of the programme. It is possible, therefore, that some users became ‘trapped’ at one level, because they forgot how to access the other levels.

To conclude, we review the study questions. The first asked whether practice with GeST would improve participants’ production of gestures. The answer here is ‘yes’, although gains were small. We also found that improvements were limited to items that featured in the programme (question 2), and did not occur when GeST was used without ongoing therapist support (question 3), although the latter finding may have been influenced by dose. More positively, gains were maintained after GeST was withdrawn (question 4). Spoken naming was not affected by GeST (question 5), despite the fact that participants repeatedly heard the spoken targets. The views of participants and their partners about GeST (questions 6 and 7) were very positive and GeST proved to be easy and enjoyable to use (question 8).

Several questions remain. One is whether the magnitude of gain can be improved, e.g. through longer practice phases. Indeed GeST did not achieve the very high dosage reported by other computer therapies (e.g. Palmer et al, 2012). The programme is currently configured to practice successive blocks of 5 items. Enabling cumulative practice, where the number of items gradually increases to 15, might induce higher usage and greater gains.

A further issue is whether accuracy of practice affects gain. In the pilot, recognition scores were only collected at level 2 of the programme, and these were found not to correlate with participants' gains. However, the small number of participants, and the fact that gesture recognition was not recorded at all levels of the programme, makes this an imperfect test.

Related to the issue of accuracy is the role of the live therapist. It may be, for example, that the SLT researcher's support in phase 1 of the pilot promoted accuracy of practice, and thereby contributed to the positive outcome. Other benefits may have related to motivation and/or navigation of the tool.

We also need to know the optimal candidates for GeST. Further research, with increased numbers of participants, could explore whether baseline factors, such as executive function or apraxia, predict usage patterns &/or gains from GeST.

The key question, however, is whether computer learning of gesture can change everyday communication skills. This, and the other remaining questions, need to be addressed in future research.

Acknowledgements

This work was funded by the Research Councils UK Digital Economy Programme (EPSRC grant EP/1001824/1). The Stroke Association was the project partner. We thank the project consultants: Emma Buswell, Justine Everson, Gerald Hartup, Philip Pepper and Carol Watson. We thank all our participants and their partners.

Footnote

[1] Some were assisted by a cunning low tech aid, involving a clipboard and a clothes peg, dreamt up by the 2nd author's Dad.

References

Archibald, L., Orange, J., & Jamieson, D. (2009) Implementation of computer-based language therapy in aphasia. *Therapy Archives of Neurological Disorders*, 2 (5), 299-311.

Boo, M. & Rose, M. (2011). The efficacy of repetition, semantic, and gesture treatments for verb retrieval and use in Broca's aphasia. *Aphasiology,* 25(2), 154-175.

Caute, A., Pring, T., Cocks, C., Cruice, M., Best, W. & Marshall, J. (in press) Enhancing communication in aphasia through gesture and naming therapy. *Journal of Speech, Language and Hearing Research.*

Cicone, M., Wapner, W., Foldi, N., Zurif, E. & Gardner, H. (1979) The relation between gesture and language in aphasic communication. *Brain and Language*, 8, 324 – 349.

Dabul, B. (2000) *Apraxia Battery for Adults*, Second ed. Austin, Tx: Pro-Ed.

Daumuller, M. & Goldenberg, G. (2010). Therapy to improve gestural expression in aphasia: A controlled clinical trial. *Clinical Rehabilitation*, 24(1), 55-65

Duffy, R. & Duffy J. (1981) Three studies of deficits in pantomimic expression and pantomimic recognition in aphasia. *Journal of Speech and Hearing Research*, 46, 70 – 84.

Duffy, J., Duffy, R. & Mercaitas, A. (1984) Comparison of the performance of a fluent and non fluent aphasic on a pantomimic referential task. *Brain and Language*, 21, 260 – 273.

Galliers, J., Wilson, S., Roper, A., Cocks, N., Marshall, J., Muscroft, S. & Pring. T. (2012) Words are not enough: empowering people with aphasia in the design process. In *Proceedings of the 12th Participatory Design Conference: Research Papers - Volume 1* (PDC '12), Kim Halskov, Heike Winschiers-Theophilus, Yanki Lee, Jesper Simonsen, and Keld Bødker (Eds.), Vol. 1. ACM, New York: NY, USA, 51-60.

Galliers, J., Wilson, S., Muscroft S., Marshall, J., Roper, A., Cocks, N. & Pring. T. (2011) Accessibility of 3D Game Environments for People with Aphasia: An Exploratory Study. *2011 ASSETS conference.*

Gloser, G., Weiner, M. & Kaplan, E. (1986) Communicative gestures in aphasia. *Brain and Language*, 27, 2, 345–359

Goldenberg, G., Hartmann, K. & Schlott, I. (2003). Defective pantomime of object use in left brain damage: Apraxia or asymbolia? *Neuropsychologia*, 41, 1565-1573

Goodwin, C. (2000) Gesture, aphasia and interaction. In: McNeill, D. (Ed) *Language and Gesture*. Cambridge: Cambridge University Press 84-98.

Hadar, U., Burnstein, A., Krauss, R. K. & Soroker, N. (1998) Ideational gestures and speech in brain-damaged subjects. *Language and Cognitive Processes*, **13**(1), 59–76.

Hogrefe, K., Ziegler, W., Weidinger, N., & Goldenberg, G. (2012). Non-verbal communication in severe aphasia: Influence of aphasia, apraxia, or semantic processing? *Cortex*, 48, 952-962.

Kemmerer, D., Chandrasekaran, B., & Tranel, D. (2007). A case of impaired verbalization but preserved gesticulation of motion events. *Cognitive Neuropsychology*, 24 (1), 70–114.

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes*, 24(2), 145-167.

Lagarno, M., Di Pietro, M. & Schnider, A. (2006) Computerised treatment of anomia in acute aphasia: Treatment intensity and training size. *Neuropsychological Rehabilitation*, 16 (6), 630 – 640.

Linebarger, M.C., McCall, D., Virata, T., & Berndt, R.S. (2007). Widening the temporal window: Processing support in the treatment of aphasic language production. *Brain and Language, 100,* 53-68.

Marangolo, P., Bonifazi, S., Tomaiuolo, F., Craighero, L., Coccia, M., Aleoe, G., Provinciali, L. & Cantagallo, A. (2010). Improving language without words: First evidence from aphasia. *Neuropsychologia*, 48, 3824 – 3833.

Marshall, J., Best, B., Cocks, N., Cruice, M., Pring, T., Bulcock, G., Creek, G., Eales, N., Lockhart Mummery, A., Matthews, N., Caute, A. (2012) Gesture and naming therapy for people with severe aphasia. *Journal of Speech, Language and Hearing Research*, 55, 726–738

Marshall, J., Atkinson, J., Smulovitch, E., Thacker, A. and Woll, B. (2004) Aphasia in a user of British Sign Language: Dissociation between Sign and Gesture. *Cognitive Neuropsychology*, 21 (5), 537 - 554.

McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press.

Palmer, R., Enderby, P., Cooper, C., Latimer, N., Julious, S., Paterson, G., Dimairo, M., Dixon, S., Mortley, J., Hilton, R., Delaney, A., & Hughes, H. (2012) Computer therapy compared with usual care for people with long-standing aphasia poststroke: a pilot randomized controlled trial. *Stroke*, 43(7), 1904-11.

Parr, S. (2007) Living with severe aphasia: tracking social exclusion. *Aphasiology*, 21, 1, 98-123

Purdy, M. & Koch, A. (2006). Prediction of strategy usage by adults with aphasia. *Aphasiology*, 20 (2-4), 337-348.

Ramsberger, G. & Marie, B. (2007) Self-Administered Cued Naming Therapy: A Single-Participant Investigation of a Computer-Based Therapy Program Replicated in Four Cases. *American Journal of Speech-Language Pathology* 16, 343-358

Raymer, A.M., Singletary, F., Rodriguez, A., Ciampitti, M., Heilman, K.M., & Rothi, L.J.G. (2006). Gesture training effects for noun and verb retrieval in aphasia. *Journal of the International Neuropsychological Society, 12*, 867-882.

Rodriguez, A., Raymer, A.M., & Rothi, L.J.G. (2006). Effects of gesture and semantic-phonologic treatments for verb retrieval. *Aphasiology, 20*, 286-297.

Rose, M. (2006). The utility of gesture treatments in aphasia. *Advances in Speech Language Pathology*, 8 (2), 92-109.

Rose, M. & Douglas, J. (2003). Limb apraxia, pantomine, and lexical gesture in aphasic speakers: Preliminary findings. *Aphasiology*, 17(5), 453-464.

Rose, M. & Douglas, J. (2008). Treating semantic deficits in aphasia with gesture and verbal methods. *Aphasiology*, 22(1), 1-22.

Skelly, M., Schinsky, L., Smith, R. & Fust, R. (1974) American Indian Sign (Amerind) as a Facilitator of Verbalization for the Oral Verbal Apraxic. *Journal of Speech and Hearing Disorders,* 39. 445-456

Swinburn, K., Porter, G. and Howard, D. (2004) *Comprehensive Aphasia Test*. New York, Psychology Press.

Whiteside, S. P., Inglis, A. L., Dyson, L., Roper, A., Harbottle, A., Ryder, J., Cowell, P. E., & Varley, R. A. (2012). Error reduction therapy in reducing struggle and grope behaviours in apraxia of speech. *Neuropsychological Rehabilitation*, 1-28.

Wilkinson, R., Beeke, S. & Maxim, J. (2010). Formulating actions and events with limited linguistic resources: Enactment and iconicity in agrammatic aphasic talk. *Research on Language and Social Interaction*, 43(1), 57–84.

Table 1: Participant Details

	Gender	Age	Handedness (post stroke hand use)	Months post stroke	Pre stroke occupation	Pre stroke computer use	Limb apraxia assessment (Dabul, 2000)
1	M	65	R (both)*	24	Buyer	Daily	Not assessed
2	F	61	R (left)	36	Administrator	Daily	90%
3	M	63	R (left)	50	Engineer	Weekly	55%
4	M	31	R (left)	24	Procurement	Daily	85%
5	F	72	L (left)	23 years	Bookmaker	None	100%
6	F	90	L (both)*	91	Not known	Occasional	70%
7	M	68	R (both)*	36	Electrician	None	90%
8	M	76	R (left)	10 years	Company Director	None	95%
9	M	57	R (left)	44	Computer consultant	Daily	95%

*all had reduced right hand function post stroke

Table 2: Mean gesture recognition and rating scores (S.Ds) across the 4 time points for each group of items.

	Time 1		Time 2		Time 3		Time 4	
	Recog Score	Rating Score	Recog Score	Rating Score	Recog Score	Rating Score	Recog Score	Rating Score
Supported GeST	2.22 (1.99)	2.36 (.87)	4.11 (2.26)	2.83 (.77)	3.11 (3.59)	2.57 (.84)	3.56 (2.74)	2.89 (.74)
Independent GeST	2.00 (2.12)	2.38 (.73)	2.44 (2.79)	2.55 (.95)	3.33 (2.87)	2.79 (.79)	3.00 (2.74)	2.64 (.98)
Familiarised	2.11 (2.37)	2.63 (.81)	3.56 (1.94)	2.89 (.88)	3.00 (2.06)	2.72 (.62)	3.56 (2.79)	2.98 (.88)
Controls	3.22 (2.73)	2.27 (.80)	2.89 (1.17)	2.43 (.52)	2.22 (1.72)	2.36 (.61)	2.67 (1.73)	2.47 (.75)

Figure 1a: The GeST Prototype

Figure 1b: GeST screen shots for level 1, 2 and 3 for the target ‘camera’

Figure 2: Mean gesture recognition scores across the 4 time points

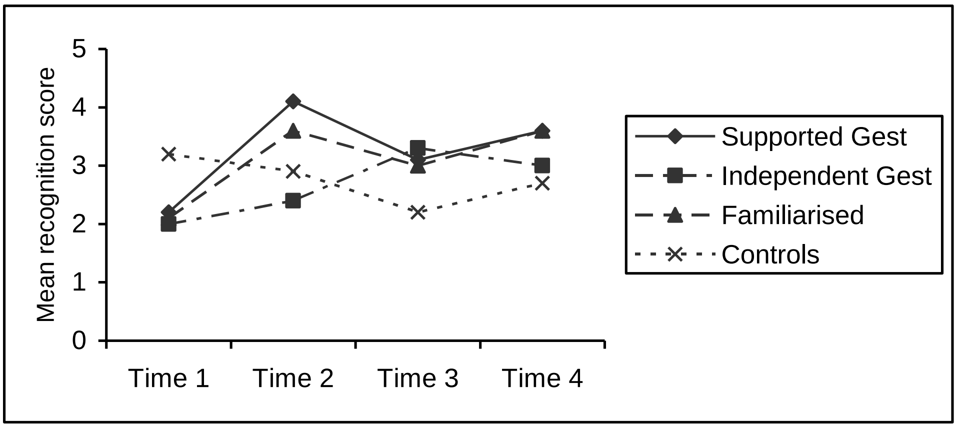


Figure 3: Mean gesture rating scores across the 4 time points

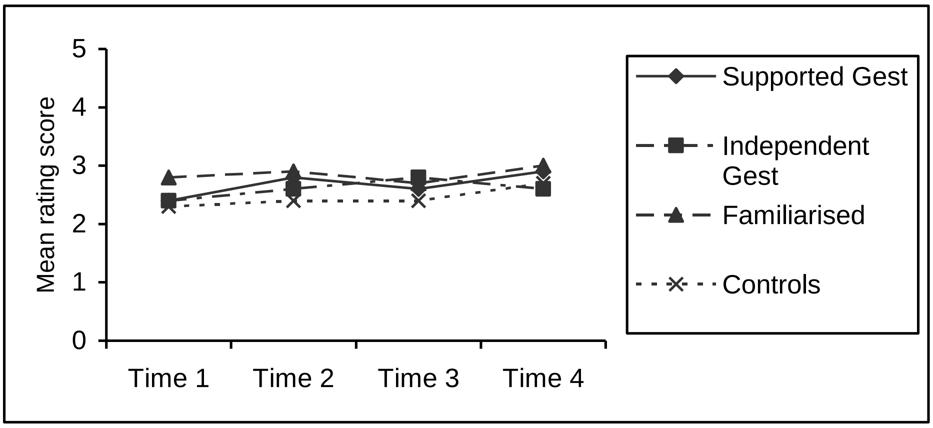


Figure 4: Mean spoken naming scores across the 4 time points

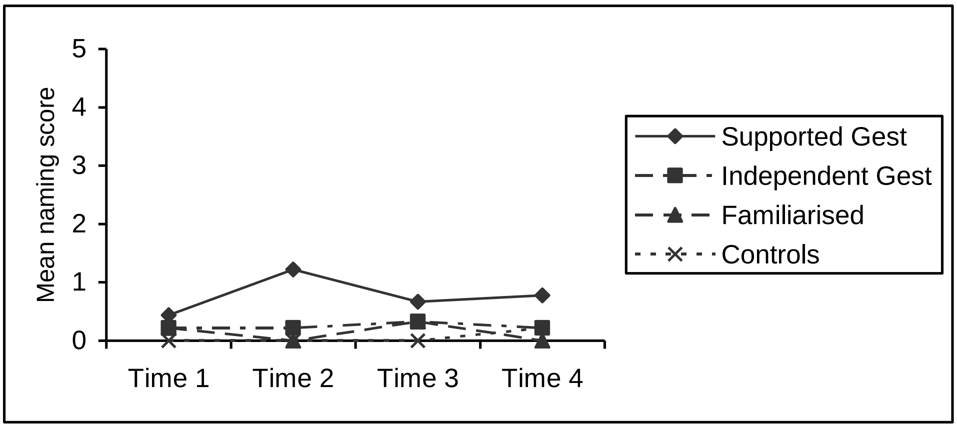


Figure 6: Number of hours used in GeST over the two phases by individual participants

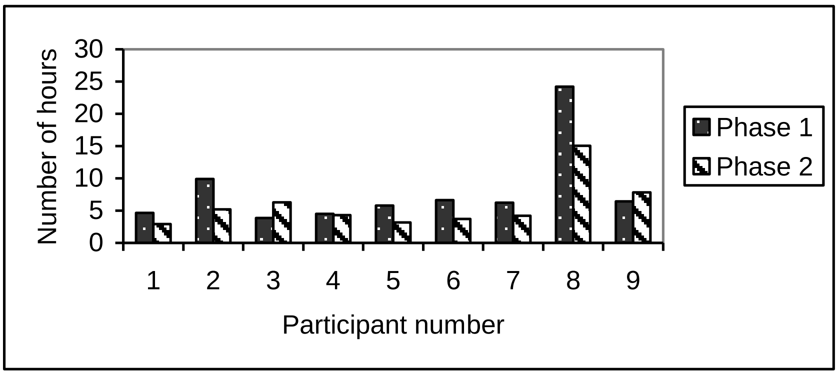


Figure 7: Number of hours used in GeST by each participant within each level (both phases combined)

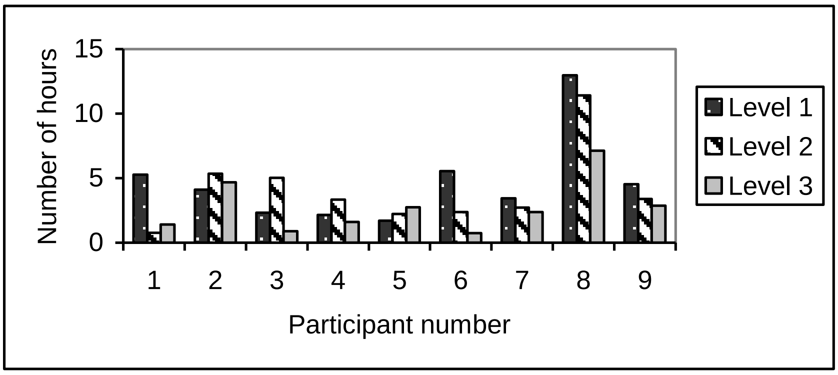
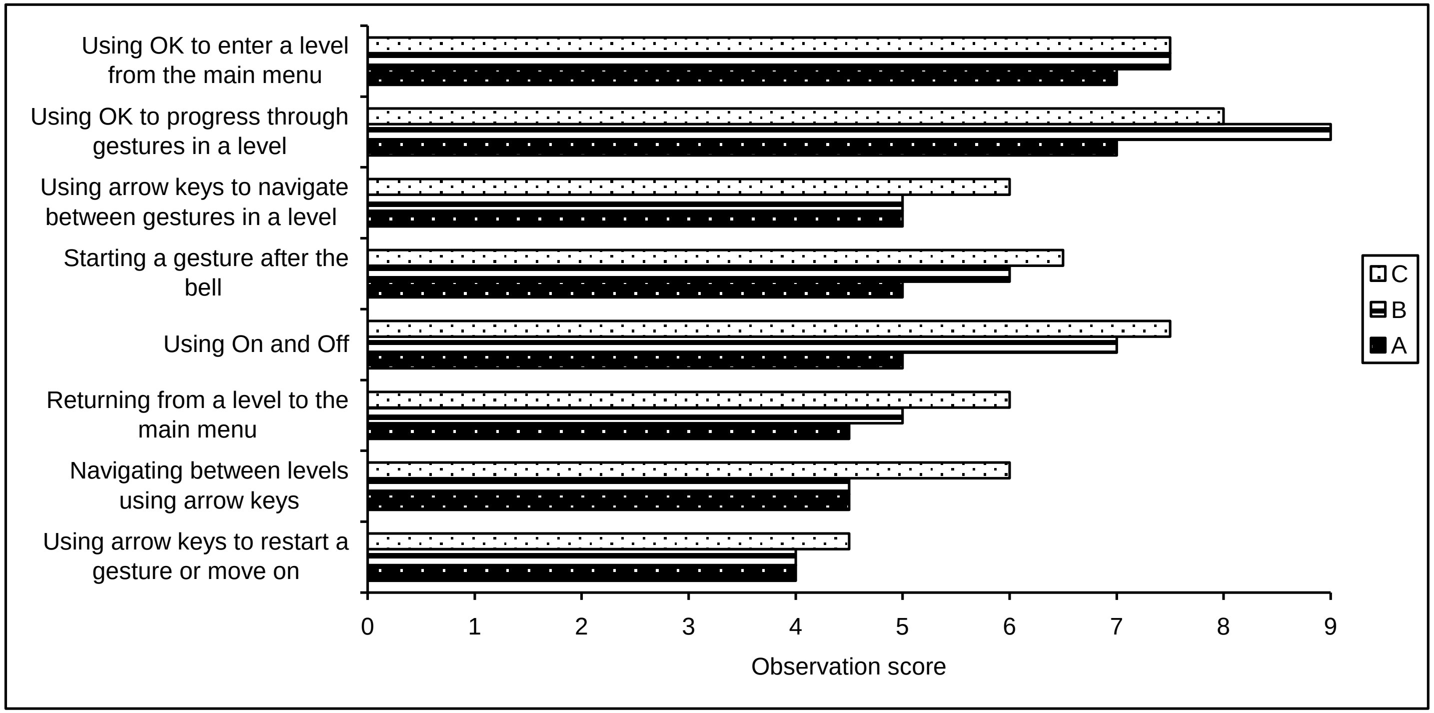


Figure 5: Total scores for structured observations of computer interaction features



Appendix: List of items included in GeST with overall gain scores (recognition score at time 4 – recognition score at time 1)

TELEPHONE	4
MONEY	4
SEWING	3
HAT	3
PIANO	3
BOY	3
BOOK	2
GLASSES	2
TEA	2
WINE	2
SPIDER	2
WALKING STICK	2
GLOVES	2
IRON	1
RAINBOW	1
DENTURES	1
STAMP	1
WAITER	1
DENTIST	1
FOOTBALL	1
WIFE	0
SCISSORS	0
BEER	0
CAR	0
FOOD	0
BED	0
CAMERA	0
TAP	0
REMOTE CONTROL	-2
UMBRELLA	-3