

ViSiCAST Milestone M6-1: Evaluation of constrained PO system

Project Number:	IST-1999-10500
Project Title:	ViSiCAST Virtual Signing: Capture, Animation, Storage and Transmission
Document Type:	Milestone Report

Milestone Number:	M6-1
Planned Date of Delivery:	Sept 2000
Actual Date of Delivery:	July 2000
Title of Milestone:	Evaluation of constrained PO system
Work-Package:	WP6 (Trials and Evaluation)
Author(s):	M. Nakisa

Abstract:

Evaluations are reported for a system – “TESSA” – developed to help sign-language communication for face-to-face transactions in the Post Office (PO). TESSA recognises what a clerk says, from a restricted list of phrases, and plays an appropriate pre-recorded phrase signed by an avatar on a screen. Six profoundly deaf people whose first language is British Sign Language (BSL) and three PO clerks took part.

The evaluations indicated that there is scope for improvement of TESSA, gave some insight into how these improvements could be achieved and provided baseline outcome measures against which improvements could be assessed. Modifications are planned for all aspects highlighted as needing improvement, including implementation of an unconstrained version, where phrases need not be repeated word for word, which should enable much more natural communication. There is a need for community evaluations to assess the views of more deaf people and further evaluations of a modified, unconstrained version of the system, eventually in a real PO setting, to establish the ultimate potential benefits of TESSA.

ViSiCAST: Evaluation of the constrained system for face-to-face communication in the Post Office - 10th July 2000

Executive summary

Evaluations are reported for a system – “TESSA” – developed to help sign-language communication for face-to-face transactions in the Post Office (PO). TESSA recognises what a clerk says, from a restricted list of phrases, and plays an appropriate pre-recorded phrase signed by an avatar on a screen. Six profoundly deaf people whose first language is British Sign Language (BSL) and three PO clerks took part. The main findings were:

- On average, 80% of the signs produced by the avatar and 61% of whole phrases were identified correctly.
- For ratings of ease of identification on a 5-point scale from 1-“Very difficult” to 5-“Very easy”, 79% of phrases were rated 3 or higher.
- For ratings of acceptability on a 3-point scale from 1-“Low” to 3-“High”, 63% of phrases were rated as 2 or 3.
- On average, the time taken to complete staged transactions was longer with TESSA than without, and the deaf participants, and to a lesser extent the clerks, rated communication with TESSA as more difficult and as less acceptable than without TESSA.
- Two of the six deaf participants said they would prefer to have the system available in the PO for use when communication became difficult. The other four said they would prefer to communicate without TESSA in its present form.
- The three deaf participants who usually experienced some worry or upset using the PO said communication with TESSA in the PO would not bother them at all.
- Aspects identified as needing improvement included facial expressions, clearer handshapes, finger configurations and lip patterns (especially for numbers and finger-spelling), the delay between spoken and signed phrases and a clearer distinction between face/hands and plain clothing.
- All clerks said they would prefer to have the system available as they thought it would make communication with deaf customers easier and more effective, though may take more time. Use of the system for multiple languages would ensure more frequent use and hence more likely use with deaf people.
- The clerks suggestions for improvement were primarily access to more phrases and an unconstrained system where phrases need not be spoken verbatim.

In conclusion:

The evaluations indicated that there is scope for improvement of TESSA, gave some insight into how these improvements could be achieved and provided baseline outcome measures against which improvements could be assessed. Modifications are planned for all aspects highlighted as needing improvement, including implementation of an unconstrained version, where phrases need not be repeated word for word, which should enable much more natural communication. There is a need for community evaluations to assess the views of more deaf people and further evaluations of a modified, unconstrained version of the system, eventually in a real PO setting, to establish the ultimate potential benefits of TESSA.

1. Introduction

Evaluations are reported for a system – “TESSA” – developed to help sign-language communication in the Post Office (PO). Six profoundly deaf people whose first language is British Sign Language (BSL) and three PO clerks took part. The evaluation comprised three main parts: identification of each phrase, performance of staged transactions and completion of questionnaires with collection of open-format feedback and views on the use of such systems.

The identification of phrases was performed to assess intelligibility of the signed phrases in TESSA. This measure would also provide a baseline for comparison of the signed content of future versions of the system. The transactions performed with and without TESSA enabled an assessment of use of the system for two-way communication in realistic PO situations. The questionnaires enabled a formal assessment of subjective views of TESSA gained throughout the trials and its likely benefit for use in the PO. These questions would also provide useful baseline measures for assessing modified versions of the system.

1.1 TESSA

TESSA recognises what a clerk says, from a restricted list of phrases, and plays an appropriate pre-recorded signed phrase on a screen. Phrases are signed by a female avatar (TESSA v.1) and can be presented with the text of the phrase on the screen¹. In the current “constrained” version of the system, the exact wording of the phrase must be spoken for recognition. Phrases were selected from transcripts of recordings of PO transactions to achieve a reasonable coverage of the most popular transactions. A set of 115 phrases were identified which were estimated to be adequate to cover about 90% of transactions performed (90% of instances of transactions rather than 90% of all possible transactions). Each phrase was recorded using motion-capture technology from signing produced by one of two profoundly deaf people whose first language is BSL. All days of the week and numbers were recorded separately so that they could be spliced within relevant phrases where appropriate.

1.2 BSL-users

Six pre-lingually profoundly deaf people whose first language is BSL took part in the evaluations. They were recruited through the deaf-UK e-mail newsgroup or through local RNID offices. Table 1 gives their biographical details. They were paid for their participation and all travel and accommodation costs reimbursed.

¹ Future versions of the system may present the text of different languages on the screen as an alternative option to presenting BSL.

BSL-user	Where live now	Where grew up	Gender	Age	Age became deaf	Other methods of communication able to use
1	Hull	Hull	M	41 - 50	Birth	SSE
2	Cambridge	Hertford	F	41 - 50	Birth	SSE
3	S.Wales	Hull	M	21 - 30	Birth	SSE; spoken English
4	Derby	Swansea	M	21 - 30	0 – 2	SSE; spoken English
5	Durham	Durham	M	21 - 30	Birth	-
6	London	Belfast	F	31 - 40	Birth	SSE

Table 1. Biographical details for the six BSL-users who took part in the evaluations. (SSE – sign supported English.)

1.3 Clerks

Three clerks were recruited by the PO to take part in the evaluations. Details of their PO working history and previous experience of communicating with deaf people in the PO are listed in Table 2.

Clerk	Gender	Current PO	Years worked as a clerk	How often serve deaf customers	How communicate with deaf customers
1	M	Trafalgar	11 - 20	1 a week	Speak; write things down;
2	F	Trafalgar	11 - 20	1 a week	Speak; write things down;
3	M	Barking	11 - 20	2 to 4 a week	Speak; write things down;

Table 2. Details of the three PO clerks who took part in the evaluations.

1.4 Protocol

The evaluations took place over three sets of two days. Two BSL-users and one clerk attended for each pair of days. The first day started with completion of the first part of the questionnaire. Each BSL-user then alternated between performing a block of transactions and identifying a block of phrases. Finally, the remainder of the questionnaires were completed at the end of the second day and any general feedback recorded. BSL – English interpreters were present throughout.

2. Phrase intelligibility

2.1 Aim

To assess the intelligibility of phrases available in TESSA and to provide a baseline measure for assessing the signed content of future versions of the system.

2.2 Method

The deaf participants were asked to identify each phrase in the system. They were presented with each signed phrase and asked to write down what they understood. They were also asked to rate each phrase for ease of identification (on a 5-point scale from 1- “Very difficult” to 5-“Very easy”) and to rate how acceptable the phrase was as an example of BSL (on a 3-point scale from 1-“Low” to 3-“High”). 133 phrases were generated from the 115 distinct phrases by incorporating appropriate days of the week and numbers to ensure that each day and each number (units and tens) was presented at least once.

Phrases were presented on a Gateway Solo 9100 laptop PC. The TESSA avatar operated at a rate of 25 frames per second. Signed phrases were presented without text. The BSL-user controlled presentation of each phrase and was allowed to repeat each phrase up to a maximum of five presentations. Phrases were presented in blocks of between 20 and 24, in groups according to broad categories, for example, postage, bill payment, amounts of money.

Accuracy of identification of phrases was assessed in two ways. First, by the accuracy of identification of whole phrases. Second, accuracy was assessed by breaking each phrase down into approximate “semantic units”. For example, “It should arrive by Tuesday but it’s not guaranteed?” would have five sign units, so “should arrive Tuesday not guaranteed” would score 5 (100%); “should arrive Tuesday” would score 3. The 133 phrases gave a total of 444 sign units. While these units were not all distinct (for example, the sign for “pound” was presented lots of times), identification of each presentation of a unit was scored separately. One experimenter judged the accuracy of responses for both measures on the basis of written responses from each BSL-user.

Once each phrase had been scored for accuracy of identification, each BSL-user was re-presented with each phrase not identified correctly along with the text of the intended phrase. With an interpreter and experimenter, they were asked to indicate whether the signs were inappropriate or whether they were just not clear.

2.3 Results

The average number of times each phrase was presented before an attempt at identification was made was 1.8. Attempts at identification were made after one presentation for the majority of phrases (51%) and required more than two presentations for 20% of phrases.

Accuracy of identification

The average accuracy of identification of whole phrases was 61% and ranged from 42% to 70% across BSL-users (Figure 1a). For the identification of sign units in phrases, average accuracy was 80% and ranged from 67% to 89% (Figure 1b).

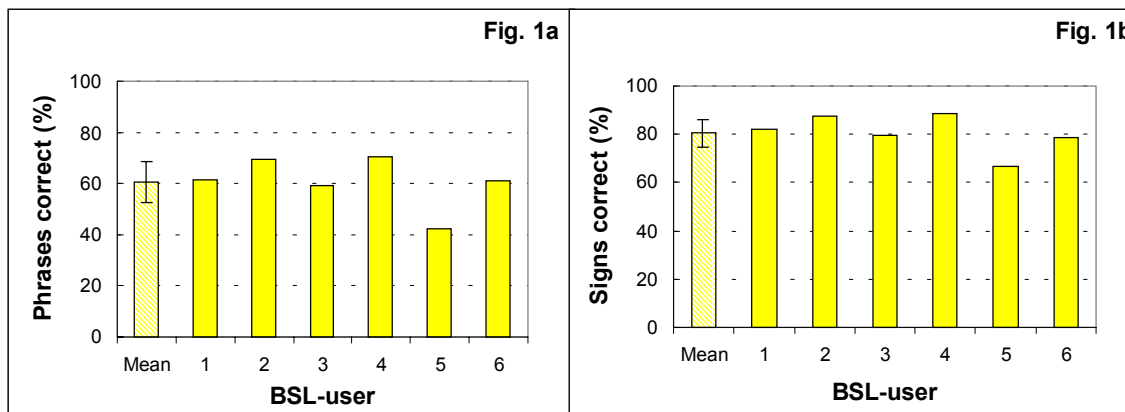


Figure 1. The average accuracy of identification of (a) whole phrases and (b) sign units in phrases achieved by each BSL-user. Error bars show the 95% confidence intervals of the overall means for the six BSL-users.

Subsequent analysis of the sign units which were wrongly identified indicated that on average 30% of errors (6% of all sign units) were due to incorrect signs and the remaining 70% (13% of all sign units) were due to unclear signing (Figure 2).

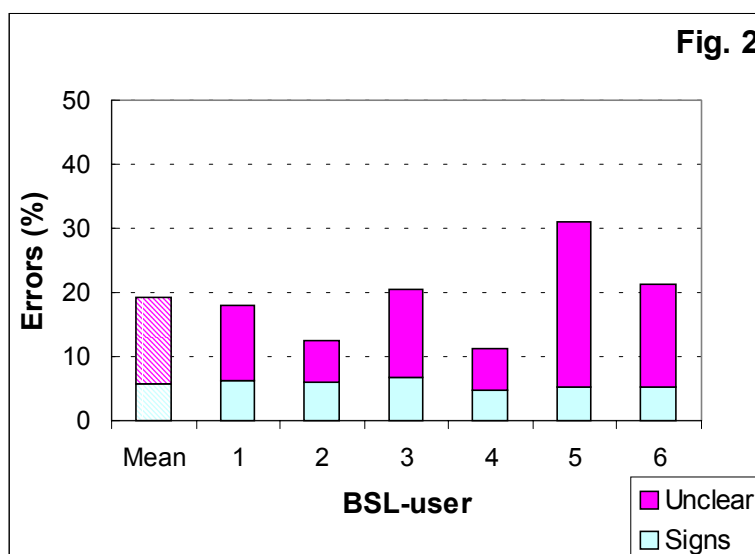


Figure 2. The percentage of identification errors made by each BSL-user categorised according to whether the error was due to an inappropriate sign (lower section of each bar) or was just unclear (upper section). Striped bars show the mean errors over all BSL-users.

Ease of identification

Table 3 shows the percentage of phrases which were rated in each category of ease of identification. The average rating was 3.6, ranging from 2.8 to 4.5 across BSL-users. The percentage of phrases which achieved each rating from each BSL-user are shown in Figure 3.

Rating of ease of identification		% of phrases
Very easy	5	31.5
	4	26.3
	3	21.6
	2	14.0
Very difficult	1	6.7

Table 3. The percentage of phrases rated in each category of ease of identification.

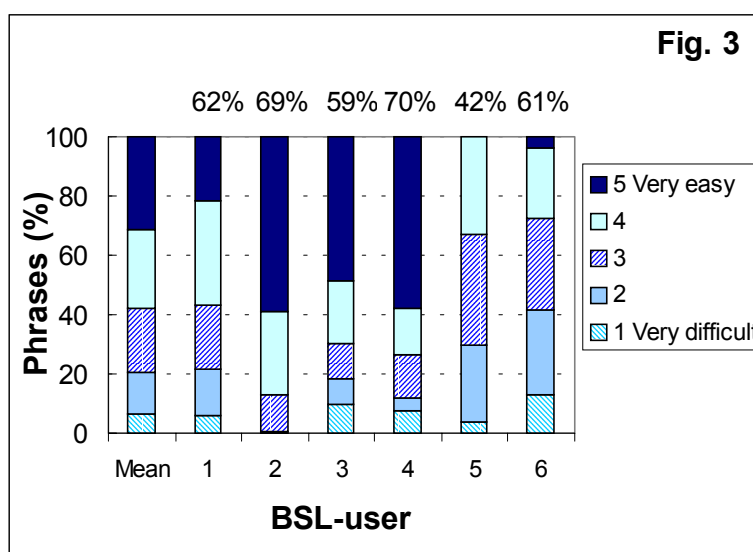


Figure 3. Percentage of ratings made by each BSL-user for ease of identification of phrases in each category on a 5-point scale ranging from 1- “Very difficult” to 5- “Very easy”. Percentage values above each bar show the accuracy of identification of whole phrases achieved by each BSL-user.

Acceptability

Table 4 shows the percentage of phrases which were rated in each category of acceptability. The average acceptability rating was 2.2 and ranged from 1.7 to 2.8. The percentage of phrases which received each rating from each BSL-user are shown in Figure 4.

Acceptability rating		% of phrases
High	3	20.2
	2	43.2
Low	1	36.6

Table 4. The percentage of phrases rated in each category of acceptability.

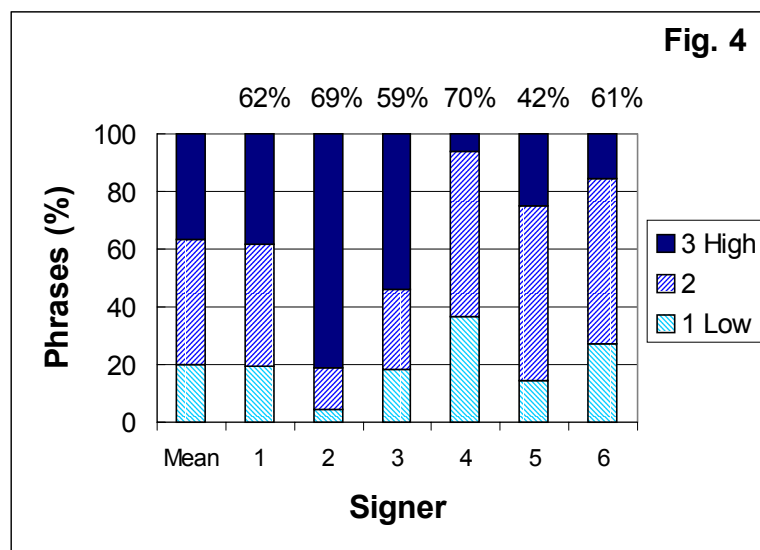


Figure 4. Percentage of ratings made by each BSL-user for acceptability of phrases in each category on a 3-point scale ranging from 1-“Low” to 3-“High”. Percentage values above each bar show the accuracy of identification of whole phrases achieved by each BSL-user.

2.4 Discussion

Accuracy of identification of the signed phrases was fairly high, with averages of 61% for whole phrases and 80% for sign units, with quite a wide range in accuracy across BSL-users (ranges of 28% and 22%, respectively). This range in accuracy suggests it is important to use many sign-language users for a true assessment of signed content of these systems. In future, it may be more appropriate to use more than six BSL-users from a range of UK regions to assess sign quality.

The majority of identification errors (70%) were due to signs being unclear rather than due to inappropriate signs. The percentage of errors for inappropriate signs did not differ greatly between subjects, with personal averages ranging from 4.9% to 6.6%. This pattern might suggest that the same signs were considered inappropriate by all BSL-users. However, inspection of the pattern of errors across BSL-users for each phrase indicated that this was not necessarily the case. Of the 46 phrases where one or more sign was considered inappropriate by any BSL-user, in 34 of these (74%) a sign was considered inappropriate by no more than two of the BSL-users. This result suggests that regional variations or differences in personal signing style may have played a role in phrase intelligibility.

While over half the phrases were considered easy to identify in the top two categories of the 5-point rating scale, and 77% in the top three categories, each BSL-user rated phrases across at least 3 categories (Figure 3). This range of rating indicates that all BSL-users perceived differences between the phrases in terms of ease of identification. Ratings of acceptability were also given across the scale with 20% of phrases rated as highly acceptable and 63% in one of the top two categories. Hence there is scope for improvement in both the subjective ease of identification and acceptability of phrases as exemplars of BSL.

As might be expected, there appears to be some relationship between average identification accuracy and ratings of ease of identification and acceptability for each

BSL-user (Figure 3). For example, BSL-users 2 and 4 who achieved the highest levels of accuracy rated more phrases as “Very easy” than BSL-user 5 who achieved the lowest level of accuracy. However, accuracy does not account for all variation in ratings. These two subjective measures therefore appear to provide additional information to numerical measures of intelligibility.

2.5 Conclusions

- Average accuracy of identification of whole phrases was 61%, ranging from 42% to 70%, and of sign units was 80%, ranging from 67% to 89% across the six BSL-users.
- On average, the majority of identification errors (70%) were due to unclear signs, rather than signs considered inappropriate. Errors made due to inappropriate signs were not always considered as such by all BSL-users hence regional variations may have played a role.
- The average ease of identification rating was 3.6 (ranging from 2.8 to 4.5) on a 5-point scale from 1-“Very difficult” to 5-“Very easy”.
- The average acceptability rating was 2.2 (ranging from 1.7 to 2.8) on a 3-point scale from 1-“Low” to 3-“High”.
- Ease of identification and acceptability ratings may be useful assessment measures in addition to numerical measures of intelligibility.

3. Transactions

3.1 Aim

To compare completion times and ease and acceptability of communication in staged transactions with and without TESSA.

3.2 Method

Each BSL-user attempted 18 transactions with a single PO clerk. Transactions were selected by the PO as those achievable with the phrases available. Of the 18 transactions, 6 were denoted simple, 6 average difficulty and 6 complex. The average difficulty and complex transactions were attempted twice by each BSL-user/clerk pair, once with an open counter and once behind a “fortress” counter (where a transparent screen separates clerk and customer). Use of different counter styles did not appear to affect performance hence results are not reported separately here. The actual complexity of the transactions varied between clerks according to the number of questions they asked and assumptions they made. Hence differences between performances for each designated category of complexity were not considered valid for analysis *post hoc*.

Half of all transactions were attempted with TESSA and half without. The phrases presented with/without TESSA were counter-balanced between BSL-users. Three practice transactions were performed with TESSA at the start of each session so that the clerk, BSL-user and interpreter could get used to using TESSA and the format of the evaluation. Transactions were performed in blocks of 6, three with TESSA and three without.

Transactions were performed using a Gateway Solo 9300 laptop PC with the avatar presented on a Packard-Bell 15” flat screen, enabling a presentation rate of at least 25 frames per second. The clerks used a hand-held Apple microphone. The speech recognition system was trained by each clerk for a quarter of an hour on the first day of testing before using the system in the trials. Sound levels were calibrated at the start of each day of testing.

The approximate time taken to successfully complete each transaction was recorded. On completion of each transaction, both BSL-users and clerks were asked to rate each transaction for ease of communication (on a 5-point scale from 1-“Very difficult” to 5-“Very easy”) and acceptability (on a 3-point scale from 1-“Low” to 3-“High”).

3.3 Results

Transaction time

On average, transactions took longer to complete with TESSA than without [$F(1,178)=61.2, p<0.001$] (Figure 5). Average times for transactions were 57s without TESSA and 112s with TESSA.

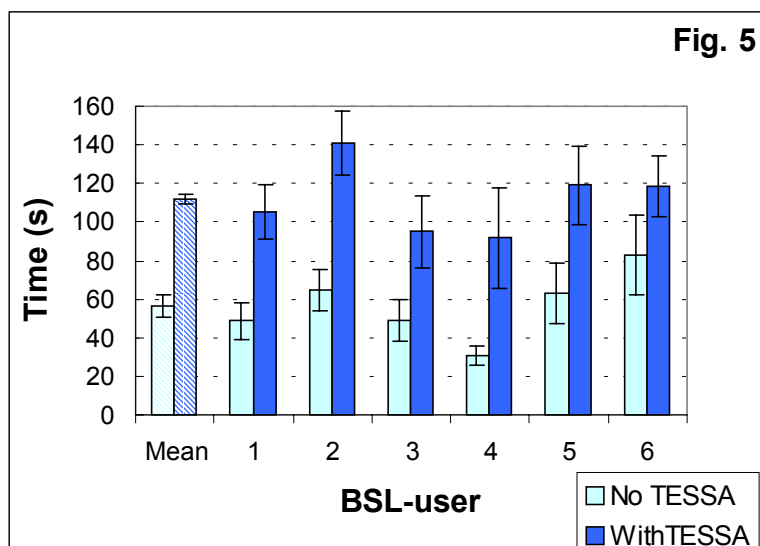


Figure 5. Average times taken for transactions without TESSA (light-coloured bars) and with TESSA (dark bars), for each BSL-user and over all BSL-users. Error bars show the 95% confidence intervals of the means.

BSL-user rating for ease of communication

On average, transactions completed with TESSA were rated as more difficult than transactions completed without TESSA [$U(1,178)=6009, p<0.001$] (Figure 6). On the 5-point scale (from 1 – “Very difficult” to 5 – “Very easy”) average ratings were 3.4 with TESSA and 4.3 without.

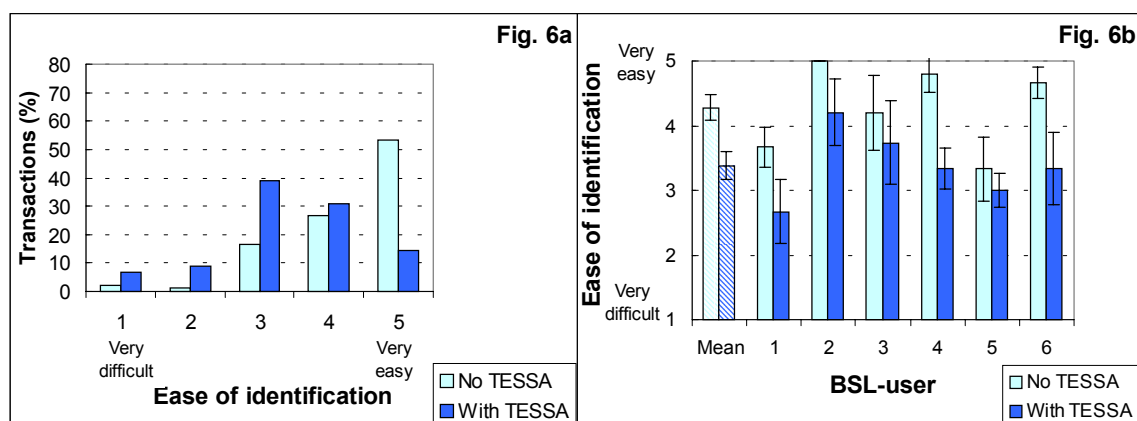


Figure 6. Ratings of ease of communication in each category on a 5-point scale from 1-“Very difficult” to 5- “Very easy” in transactions without TESSA (light bars) and with TESSA (dark bars). (a) Percentage of transactions rated in each category. (b) Average ratings for each BSL-user and over all BSL-users. Error bars show the 95% confidence intervals of the means.

BSL-user rating for acceptability

On average, communication in transactions completed with TESSA was rated as less acceptable than in transactions completed without TESSA [$U(1,178)=6025, p<0.001$] (Figure 7). On the 3-point scale (from 1 – “Low” to 3 – “High”) average ratings were 1.9 with TESSA and 2.6 without.

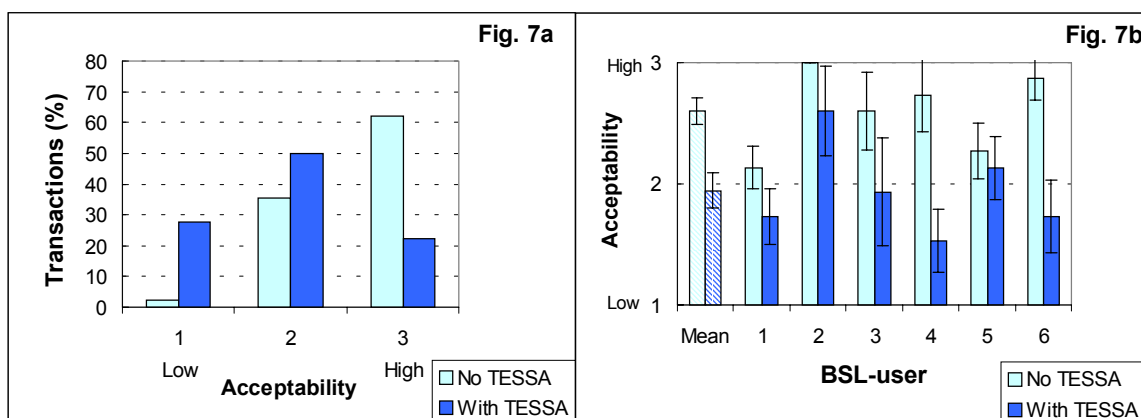


Figure 7. Ratings by BSL-users for acceptability on a 3-point scale from 1-“Low” to 3- “High” of transactions without TESSA (light bars) and with TESSA (dark bars). (a) Percentage of transactions rated in each category. (b) Average ratings for each BSL-user and over all BSL-users. Error bars show the 95% confidence intervals of the means.

Clerk rating for ease of communication

On average, clerks rated communication in transactions completed with TESSA as more difficult than transactions completed without Tessa [$U(1,178)=5232$, $p<0.001$] (Figure 8). On the 5-point scale (from 1 – “Very difficult” to 5 – “Very easy”) average ratings were 4.0 with Tessa and 4.4 without.

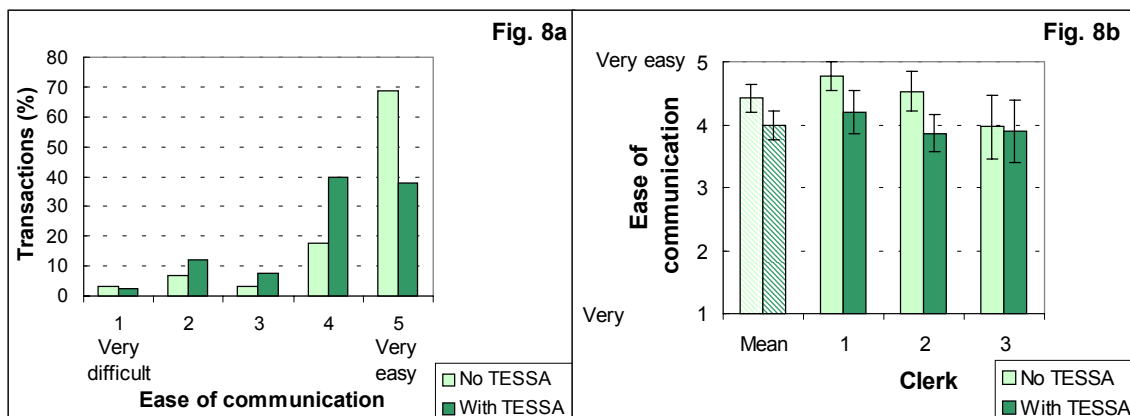


Figure 8. Rating of communication by clerks on a 5-point scale from 1-“Very difficult” to 5- “Very easy” for transactions without TESSA (light bars) and with TESSA (dark bars). (a) Percentage of transactions rated in each category. (b) Average ratings from each clerk and the overall mean rating. Error bars show the 95% confidence intervals of the means.

Clerk rating for acceptability

On average, clerks rated communication in transactions completed with Tessa as less acceptable than in transactions completed without Tessa [$F(1,178)=3.89$, $p<0.05$] (Figure 9). On a 3-point scale (from 1 –“Low” to 3 –“High”) average ratings were 2.5 with Tessa and 2.6 without.

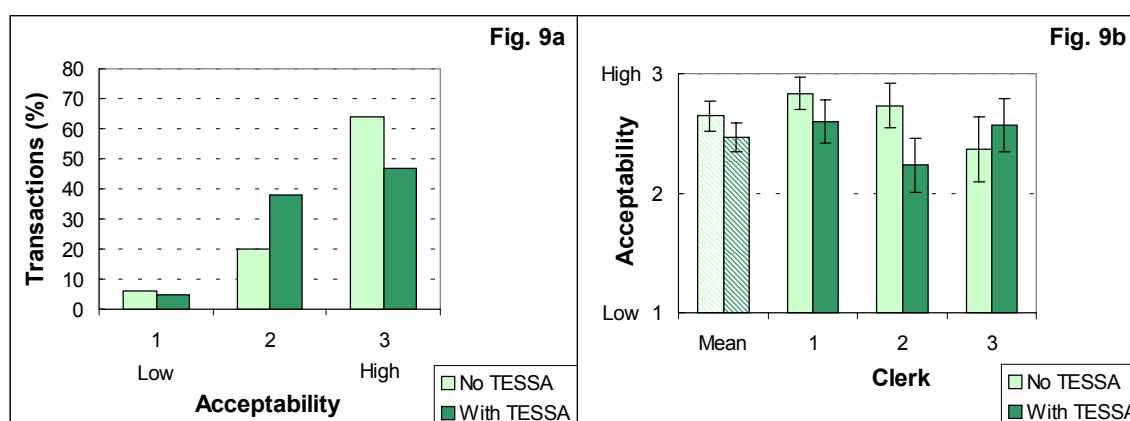


Figure 9. Ratings by clerks for acceptability on a 3-point scale from 1-“Low” to 3- “High” of transactions without TESSA (light bars) and with TESSA (dark bars). (a) Percentage of transactions rated in each category. (b) Average ratings for each clerk and over all clerks. Error bars show the 95% confidence intervals of the means.

Discussion

Compared to transactions without TESSA, transactions performed with TESSA on average took twice as long to complete and the deaf participants, and to a lesser extent the clerks, rated communication as more difficult and less acceptable. The main reason most likely to have contributed to these effects was the somewhat disjointed communication with TESSA. As expected, it took the clerks some time to learn which phrases were available and to locate the phrase they wanted on a list so they could read it out word for word. The clerks had only about an hour of practice using the system before the trials. These difficulties are likely to decrease substantially with training and experience on the system, and moreover with use of the next, unconstrained version of the system where phrases do not need to be repeated verbatim.

Additional factors may have contributed to the longer transaction times and poorer ratings with TESSA. First, the list of phrases were selected for use in the system as those most commonly used in the PO. These phrases also tended to be those used for the more simple PO transactions, for example, buying stamps, cashing a giro cheque or claiming a pension payment. As transactions used in this evaluation were limited by the phrases available, they also tended to be fairly simple or were simplified. This was confirmed by the PO staff who selected the transactions and the clerks who often said they would usually ask more questions for specific transactions but these were not available in TESSA. The transactions used in the trials therefore tended to represent situations in which communication was fairly easy without TESSA.

Second, the deaf participants were all fairly good communicators and all had reasonable written skills. Hence they were able to complete the simple transactions, by lip-reading/speaking and writing notes or asking the clerk to write things down where necessary. This is a consequence of the type of people who would be prepared to attend two days of testing away from home, the recruitment process (through e-

mail and professional connections) and also the necessary use of textphone, fax and e-mail for the logistics of arranging the trials.

Third, the clerks either were deaf aware or soon became deaf aware as a result of spending two days with the profoundly-deaf participants. Communication without TESSA was fairly easy as they used good eye contact, spoke clearly and were prepared to write things down if they were not understood.

Fourth, there was a delay of a few seconds between recognition of the spoken phrase and the signing of the phrase. Not only did this absolute delay add to overall transaction time but the delay often resulted in loss of attention and the need for the sign to be repeated or the clerk to repeat the phrase.

3.5 Conclusions

- Transaction times were on average twice as long with TESSA (average of 112s) than without (average of 57s).
- Compared to communication without, the BSL-users rated communication with TESSA as more difficult by an average of 1 point on a 5-point scale and as less acceptable by an average of 0.7 point on a 3-point scale.
- Compared to communication without, the clerks rated communication with TESSA as more difficult by an average of 0.4 point on a 5-point scale and as less acceptable by an average of about 0.1 point on a 3-point scale.
- The most likely reasons for the longer transaction times and poorer ratings with TESSA include that:
 - the clerks had little experience of using the system and so were unfamiliar with the phrases which needed to be spoken word for word;
 - the transactions were simple and so fairly easy without TESSA;
 - the deaf participants were fairly good communicators with reasonable written skills;
 - the clerks were deaf aware;
 - there was a noticeable delay between spoken and signed phrases.

4. Questionnaires

4.1 Aim

To obtain subjective views of previous experiences of communicating in the PO (BSL-users) / communicating with deaf people in the PO (clerks) and ease of communication using TESSA for face-to-face transactions in the PO, in addition to general views on the use of avatars for sign-language communication.

4.2 Method

The experimenter administered the questionnaires (Appendices 1 and 2) to both BSL-users and clerks. Questions 1 to 7 for the BSL-users and questions 1 to 5 for the clerks were asked at the start of the evaluation. The remaining questions were asked at the end when any general feedback was also recorded. The BSL-users' questionnaire included use of three visual analogue scales where BSL-users were asked to rate clarity of signing, acceptability of the avatar as a signer of BSL and the avatar appearance. This was achieved by marking on a line with "Not at all" (0%) at one end and "Totally" (100%) at the other end. Location of marked points on the line yielded a score between 0% and 100% on each scale for each BSL-user.

4.3 Results

BSL-users

All BSL-users said that if they needed something from the PO they would usually go themselves and would attempt to communicate by lip-reading/speaking and gesturing. Five BSL-users said they would also write things down and ask the clerk to write things down if needed. BSL-user 5 said he would take someone to interpret if communication became difficult and usually refused to use written communication (although he did use written communication in some of the trials without TESSA in the evaluation).

Results from the three questions asking about ease of communication in the PO, previously, in the trials with TESSA and anticipated in everyday life with TESSA are shown for each of the six BSL-users in Table 5. For the trials, ratings stayed the same or were worse for all but BSL-user 1 who said communication was "Very difficult" previously but "Manageable" with TESSA in the trials and anticipated to be "Fairly easy" with TESSA for everyday life.

When asked about the relative ease of communication with and without TESSA, in the trials and anticipated in everyday life (Table 6), four and five BSL-users (respectively) said it would be "Slightly worse" or "Much worse" with TESSA. BSL-user 1, who said that communication in the PO was usually "Very difficult", said it would be "Slightly easier" with TESSA (for both questions).

When asked about how much communication in the PO usually bothered them, three of the BSL-users said "A little" or "Not at all" (Table 7). When asked about communication with TESSA in everyday life, five of the BSL-users said this

would bother them “A little” or “Not at all”. The other BSL-user was not bothered at all previously but anticipated that using TESSA in everyday life would bother him “Quite a lot”.

Question	Very easy	Fairly easy	Manageable	Slightly difficult	Very difficult
11 How easy do you usually find communication in the Post Office?	2		3 4 6		1 5
12 How easy did you find communication using TESSA?			1 2 3 4	6	5
14 In everyday life, how easy do you think communication would be using TESSA?		1 2	6	4	3 5

Table 5. Responses made by each BSL-user to the three questions about ease of communication in the PO: previously, in the trials with TESSA and anticipated in everyday life with TESSA. Each number represents the responses from one BSL-user.

Question	Much easier	Slightly easier	No difference	Slightly worse	Much worse
13 Compared to communication without, do you think TESSA made communication:		1	3	2 4	5 6
15 In everyday life, do you think that using TESSA in the Post Office would make communication:		1	2	3 4 5	6

Table 6. Responses made by each BSL-user to the two questions about comparing communication in the PO with and without TESSA: in the trials and anticipated in everyday life. Each number represents the responses from one BSL-user.

Question	Very much	Quite a lot	Some	A little	Not at all
9 In everyday life, how much does communication in the Post Office upset, annoy or worry you?	1	4	3	6	2 5
16 In everyday life, how much would communication using TESSA in the Post Office upset, annoy or worry you?		5		6	1 2 3 4

Table 7. Responses made by each BSL-user to the two questions about how much communication in the PO bothered them, previously and anticipated with TESSA in everyday life. Each number represents the responses from one BSL-user.

When asked for a preference, four BSL-users said they would prefer to communicate without TESSA and two preferred with, as an option if needed. All BSL-users said they would prefer TESSA with both BSL and text, rather than just BSL or just text alone.

Using visual analogue scales (Figure 10), the mean rating for clarity of signing was 28% and ranged from 12% to 46% across BSL-users. For acceptability of the avatar as a signer of BSL, the mean rating was 32%, ranging from 3% to 65%. The mean rating for appearance of the avatar was 30%, ranging from 2% to 66%.

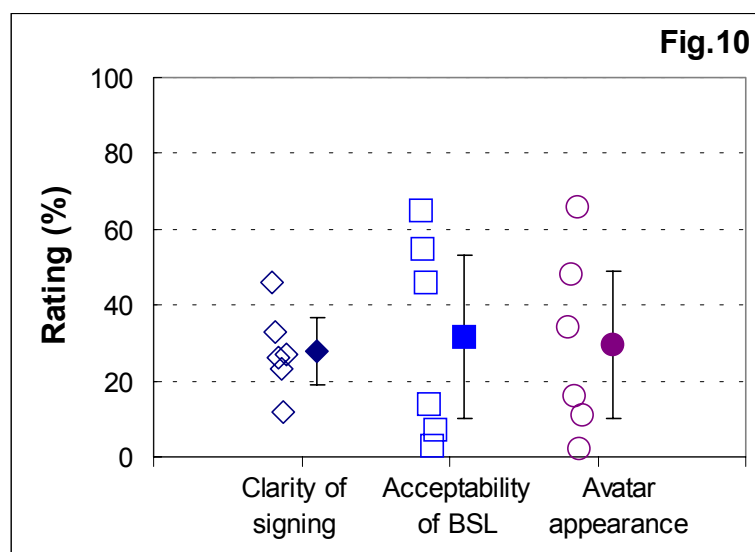


Figure 10. Ratings on visual analogue scales for clarity of signing (Qu.19), acceptability of avatar as a signer of BSL (Qu.20) and avatar appearance (Qu.21). Open symbols show ratings for each BSL-user with filled symbols showing the average rating. Error bars show the 95% confidence intervals of the means.

Clerks

All clerks rated usual communication with deaf customers as “Fairly easy” and communication with TESSA as “Fairly easy” or “Very easy” (Table 8). All clerks said communication was “Slightly easier” or “Much easier” with TESSA than without, and that in everyday life they anticipated that communication would be “Much easier” with TESSA (Table 9). All clerks said that they would prefer to have TESSA available as an option to use when communication became difficult, even though they all thought transactions would take “Slightly longer” with TESSA.

Question	Very easy	Fairly easy	Manageable	Slightly difficult	Very difficult
5 How easy do you usually find communication with deaf customers?		1 2 3			
6 How easy did you find communication using TESSA?	3	1 2			
8 In everyday life, how easy do you think communication would be using TESSA?	2 3	1			

Table 8. Responses made by each clerk to the three questions about ease of communication with deaf customers: previously, in the trials with TESSA and anticipated in everyday life with TESSA. Each number represents the responses from one clerk.

Question	Much easier	Slightly easier	No difference	Slightly worse	Much worse
7 Compared to communication without, do you think TESSA made communication:	3	1 2			
9 In everyday life, do you think that using TESSA in the Post Office would make communication:	1 2 3				

Table 9. Responses made by each BSL-user to the two questions about comparing communication in the Post Office with and without TESSA, in the trials and anticipated in everyday life. Each number represents the responses from one clerk.

4.4 Discussion and general comments

Of the six deaf participants, one person said that communication would be easier with TESSA and two BSL-users said they would prefer to communicate with TESSA in the PO, as an option in case communication became difficult. The three deaf participants who said that communication in the PO usually upset or worried them, said they thought using TESSA in the PO would not bother them at all. While this represents positive feedback from some BSL-users, the fact that these responses were not more generally positive does not seem unreasonable. These questions were asked about the first version of TESSA to be evaluated by deaf people and on the basis of use during the trials by clerks with little previous experience of using the system, where communication with TESSA was somewhat lengthy and disjointed.

Scores on the visual analogue scales showed a wide range of responses between BSL-users for ratings of clarity of signing, acceptability of the avatar as a signer of BSL and the avatar's appearance. These scales proved easy to use as the BSL-users responded more easily than to the previous questions for which it was

often difficult to obtain a categorical response. These scales therefore are likely to be useful outcome measures for evaluating future signing avatars and versions of the system. Scores on the scales were all under about 65, hence there is much scope for improvement. The deaf participants provided a lot of constructive feedback about how TESSA could be improved, described below.

Signing

There were a number of comments about the quality of signing. First, the numbers were not clear; this was stated as a major contributory factor to the low scores for the clarity of signing. While there is much regional variation in signs for numbers, it was generally felt that as long as one system was selected, used consistently and the handshapes and lip patterns were very clear then this would not be a great problem. Having the text on the screen would also help to clarify any numerical ambiguities.

Second, lip patterns should be clearer for all signs. As for the numbers, clear lip patterns would help distinguish any unclear, ambiguous or regionally-different signs.

Third, all BSL-users mentioned that it was vital that the avatar should have more facial expression. Facial expression is a vital element of BSL and must therefore be incorporated in any avatar which attempts to produce BSL. The lack of facial expression was stated as a primary factor in low ratings of acceptability of the avatar as a signer of BSL.

Fourth, it was felt that the hands and finger configurations should be generally clearer to give more precise signing and finger-spelling.

Some people felt that SSE should be available as an option, though none of the deaf participants said they would use it, preferring to use BSL. They also suggested it was important to involve older people in the evaluations as a lot of them used a more SSE-like signing method or used a lot of finger-spelling. All deaf participants said they would prefer to have both BSL and text, rather than just BSL or text alone, as they felt this would likely benefit more people and help clarify any signs which were not clear.

The avatar's appearance

Various comments were made about the appearance of TESSA. The primary concern was that the clothing should be plain as any patterns (i.e. the collar on the avatar's shirt) distract from the signing. It was generally felt that signing would be clearer if there was greater distinction between face/hand colour and clothing. For example, a paler skin tone with darker clothes or darker skin with lighter clothing. The bright blue background was also considered distracting and should be made less bright. The "jittering" thumbs in rest mode were also considered a little distracting.

The system

The delay between the clerk saying a phrase and the signed phrase was unacceptably long. Some people were also bothered by having to look away from the clerk to the third-party of the screen. It was suggested that this could be improved by locating the screen nearer to the clerk so that it would be easier to maintain eye contact between the clerk and customer.

There were many comments that wherever the system was eventually used, careful consideration must be given to the location of the screen and the lighting. There were concerns that in the PO, reflections of people in a queue behind would distract from the signing.

Use of avatars for sign language

For simple PO transactions in these trials, the deaf participants did not find a lot of benefit from using the current version of TESSA. Most of the deaf participants mentioned that deaf awareness training for all clerks would currently be a more effective means of improving communication in the PO. Nonetheless, it may be the case that even for simple PO transactions some deaf people could benefit, such as those with poorer general communication skills and poor writing/reading skills. Views of a wider range of deaf people will be sought in the community evaluations. It is also likely that as the system is improved, enabling more natural communication, and clerks become more proficient users of the system, then using TESSA could be easier and more acceptable than having to write things down.

Several of the deaf participants said that TESSA might be more helpful for situations requiring more complex communication. For example, completing complex forms for benefits or passports is a situation which often poses problems for deaf people whose first language is BSL.

Two of the deaf participants said they would definitely prefer real BSL-users to an avatar in all situations, including television. Unless the avatar was much more realistic than TESSA v.1, with facial expressions and much clearer signing then it could never be acceptable.

One of the deaf participants said that such a system for face-to-face communication should not be distributed for general use until a sign-recognition component is in place. In its current form, it is very much a hearing solution; it makes life much easier for the clerk but the deaf person is still left trying to communicate their intentions to the clerk and has to write things down as before.

Clerks

The clerks were generally positive about TESSA; they all preferred to have the system available and all thought that communication with deaf customers would be easier, even though they thought that using TESSA would make transactions longer. Having the option to use the system would mean they were more likely to ask all the questions they would normally ask which are often omitted currently when communication becomes difficult.

The clerks gave a number of suggestions when asked how the system could be improved. They would like more phrases and greater flexibility so the phrases did not have to be spoken verbatim for recognition. The delay between saying the phrase and the signing was also unacceptable. In terms of everyday use of the system, they felt that one or two clerks need to be specially trained at each PO. They were concerned that if the system was not used often then it might never be used. Use of the system for multiple languages would generate more frequent use and hence more likely use for deaf people. In addition, they thought that at least those to be trained to use the system should also receive deaf awareness training.

Minor comments included use of a microphone on a head set if the system was to be used all day and perhaps a switch on the microphone to initiate speech recognition rather than using the keyboard space bar. Also, it would be helpful if the system could be incorporated into their existing PC systems to conserve limited desk space.

4.5 Conclusions

- Two of the six deaf participants said they would prefer to have TESSA available in the PO for use when communication became difficult. The other four said they would prefer to communicate without TESSA in its present form.
- One of the deaf participants said that communication in the PO would be easier with TESSA..
- The three deaf participants who said that communication in the PO usually upset or worried them, said they thought using TESSA in the PO would not bother them at all.
- Evaluations of the unconstrained, improved version of the system are needed to establish possible benefits for face-to-face communication in the PO and community evaluations are needed to assess the views of a wider range of deaf people.
- Ratings of clarity of signing, acceptability of TESSA as a signer of BSL and TESSA's appearance ranged between BSL-users, but with scores under 65 there is scope for improvement in all three areas.
- Suggestions for improvement included:
 - Facial expressions.
 - Clearer handshapes, finger configurations and lip patterns, especially for numbers and finger-spelling.
 - Reduced delay between spoken phrase and signed phrase.
 - Improved appearance of avatar with clearer distinction between face/hands and clothing, which should be plain.

- All BSL-users said they would prefer to see both BSL and text rather than just BSL or just text. They also thought that SSE should be available as an option though they would not personally use it.
- Comments on use of avatars for signing in general included that the deaf participants find avatars more useful for more complex communication needs, e.g. explaining benefits forms.
- All clerks said they would prefer to have the system available as they thought it would make communication with deaf customers easier and more effective. Use of the system for multiple languages would ensure more frequent use and hence more likely use with deaf people.
- The clerks would like more phrases and an unconstrained system where phrases need not be spoken verbatim.

5. General conclusions

The evaluations have indicated that there is much scope for improvement of TESSA, gave some insight into how these improvements could be achieved and provided baseline outcome measures against which improvements could be assessed. The majority of aspects identified for improvement are planned for further development within the ViSiCAST project. Primarily, the development of an unconstrained version, where phrases need not be repeated word for word, would enable much more natural communication and should greatly reduce the time taken for transactions, so is also likely to be more acceptable to both deaf customers and clerks. Other aspects to be explored include research into facial modelling which should improve avatar facial expressions and lip patterns. New data gloves are also being used to improve recording of finger movements and handshapes. New models of the avatar and clothing will also take account of the comments made by the deaf participants.

Community evaluations are needed to assess the views of more deaf people and further evaluations of a modified, unconstrained version of the system are needed to establish the ultimate potential benefits of TESSA.