

# Experimental Assessment of the Effectiveness of Synthetic Personae for Multi-Modal E-Retail Applications

Helen McBreen<sup>†</sup>

Paul Shade<sup>†</sup>

Mervyn Jack<sup>†</sup>

Peter Wyard<sup>‡</sup>

<sup>†</sup>Centre for Communication Interface Research

<sup>‡</sup>BT Laboratories

University of Edinburgh

Adastral Park

80 South Bridge, EH1 1HN, Scotland, UK

Ipswich, England, UK

+44 131 650 2779

+44 1473 640 192

helen@ccir.ed.ac.uk

paul@ccir.ed.ac.uk

maj@ccir.ed.ac.uk

peter.wyrd@bt.com

## ABSTRACT

This paper details results of an experiment to empirically evaluate the effectiveness and user acceptability of human-like synthetic agents in a multi-modal electronic retail scenario. The synthetic personae played the roles of interactive conversational sales assistants. The range of life-like personae differed with respect to gender and technology. Participants took part in the controlled experiment, which involved them eavesdropping on spoken dialogues between a customer and each of the synthetic personae. They also completed questionnaires and took part in a debriefing interview designed to elicit information relating to the effectiveness, believability and perceived quality of each of the personae.

Results show that participants expected a high level of realistic and human-like verbal and non-verbal communicative behaviour in the synthetic personae. This was demonstrated in the strong preference for personae that exhibited natural facial expressions, gestures and emotions. It was also found that disembodied voices were significantly preferred to many of the personae. In addition, results show participants had significantly different attitudes to the voices of the personae.

## Keywords

*Synthetic personae, anthropomorphic, verbal and non-verbal communication, virtual conversation.*

## 1. INTRODUCTION

Synthetic personae are now a feature of many graphical user interfaces including enhanced multimedia presentations [1] and educational spoken dialogue interfaces featuring conversational characters [8]. Although such personae have been technically developed for applications little is known about user attitudes towards them [12,13]. As a result, this paper describes the first in a series of experiments to investigate user attitudes to different types of synthetic personae in electronic retail environments. The long-term aim of this work is to create a body of experimental

data to provide a set of design guidelines for the creation of effective synthetic personae to be used in intelligent multimedia and multi-modal applications involving automatic speech recognition and automatic speech generation technologies.

In such graphical interfaces, the user must be convinced that the synthetic persona can support a face-to-virtual-face conversation. Given the current limited knowledge about the naturalness of such conversations there is a risk associated with the introduction of anthropomorphic characters into GUI's. If the synthetic persona exhibits visually sophisticated communicative behaviour but cannot support the actual dialogue, the interaction between the user and the synthetic persona may collapse. Interface designers need to be aware that user expectations should be determined by the capability of the spoken language interface and this is the reason why fundamental research into understanding the relationship between user attitudes, user expectations and system behaviour is necessary.

It is important firstly to establish the types of synthetic personae with which users would like to interact, if any, and then provide the characters with the necessary verbal and non-verbal communicative behaviour to enhance face-to-virtual-face conversation. In summary, although enhancing interfaces with synthetic personae is no longer a novel idea, theoretical research into the design of such personae is in its infancy. The focus of such interface design should now be "research that can contribute to advancing innovative concepts and that can promote better understanding of what technology works well to make interfaces more usable, useful and accessible", [2].

## 2. EXPERIMENT PROCEDURE

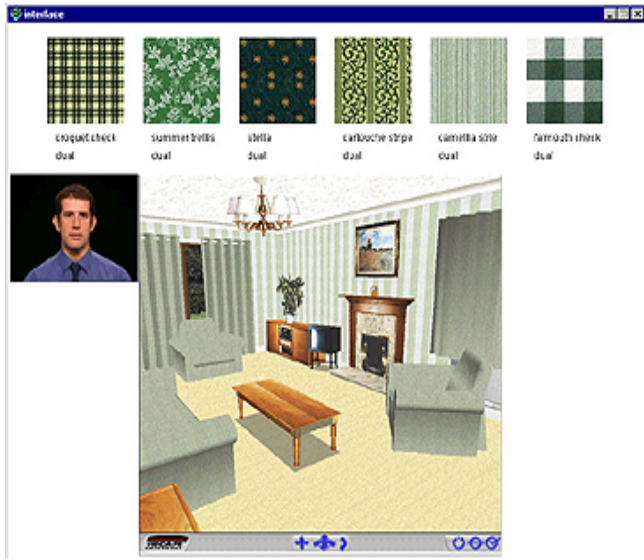
This experiment aimed to assess a wide range of usability attributes of ten synthetic personae in the context of a multi-modal electronic retail application by having participants 'eavesdrop' on brief dialogues between a customer (represented by a disembodied voice) and a synthetic persona. The passive methodology used to assess the personae was extremely practical, as it avoided the complex technological issues involved in creating a fully functional interactive application with a range of personae, but it still allowed a full evaluation of each of them.

The GUI for the retail application (Figure 1) was created in the style of MUESLI [14]. The main window was a 3D view of a living room complete with furniture. Immediately above was a row of fabric and wallpaper samples that could be selected in order to 'decorate' the walls, sofa, chairs and curtains. The

---

**Note:** The primary author of this paper is a PhD student

synthetic persona was displayed in its own window (top left). The dialogue illustrated the ‘customer’ conversing with the persona to select colours and patterns in order to decorate the room.



**Figure 1: The E-Retail Application Interface**

The ten personae differed with respect to the technology (T) used to implement them and gender. Male and female versions of five contrasting technologies were used as follows:

**T1:** Videos of human (male and female) sales assistants. The female voice soundtrack was used for the other female personae; the male voice soundtrack was used for the other male personae.

**T2:** 3D talking heads (modelled on T1, T3 and T4) with lip-synchronisation to the original male and female voice soundtracks.

**T3:** Still frames taken from the videos of the sales assistants with added graphic lip movement to match the original voice soundtracks.

**T4:** Still frames taken from the videos of the sales assistants (as in T3) *without* graphic lip movement.

**T5:** Disembodied male and female voices (original voice soundtracks).

The dependent variables in the experiment were the responses to the individual items in the questionnaires, listed in Table 2 and responses given during a semi-structured interview, including an overall rating of each persona. The independent variables were persona gender and technology. Effects of between-subject variables of age, gender and experiment supervisor were also investigated. A total of 32 participants took part in the experiment, distributed according to gender and age as shown in Table 1.

| Participant Age Group | No. of Males | No. of Females | Total |
|-----------------------|--------------|----------------|-------|
| 18-35                 | 8            | 5              | 13    |
| 36-49                 | 1            | 5              | 6     |
| 50+                   | 7            | 6              | 13    |
|                       | 16           | 16             | 32    |

**Table 1: Analysis of Participants by Gender and Age Group.**

The experimental procedure required participants first of all to read a brief explanation of the purpose of the experiment after which they were also primed verbally by the experiment supervisor. They then viewed ten 2-minute videos (created using Macromedia Director 6.5 and presented in randomised order on a Pentium II PC), showing the dialogue between the ‘customer’ and one of the synthetic personae.

**Customer:** *I'd like to plan a make over for my sitting room.*

**Assistant:** *Good, what would you like to see first?*

**Customer:** *Can you show me some green fabrics for the sofa?*

**Assistant:** *There are more than forty green fabrics, here is a selection.*

**Customer:** *Try the Cartouche Stripe please.*

**Assistant:** *Would you like to see it on the chairs as well?*

**Customer:** *Yes, ok.*

**Figure 2: Section of the Dialogue**

The same dialogue with the same male and female voice recordings was used throughout. After listening to each dialogue, participants completed a Likert questionnaire [5,7], formatted as shown in Figure 3. Within the questionnaire, statements were balanced for polarity (equal number of positively and negatively worded stimulus statements).

**I liked the appearance of the assistant.**

|                          |                          |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Strongly Agree           | Agree                    | Slightly Agree           | Neutral                  | Slightly Disagree        | Disagree                 | Strongly Disagree        |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

**Figure 3: Example of a Likert Questionnaire Item**

Not all dimensions of interest were relevant to all the synthetic personae, therefore three different questionnaires were used. A total of eight statements were relevant to all the technologies and were therefore included in all questionnaires. Other statements relating to appearance (not relevant to the disembodied voice condition T5) and lip-movement were included as appropriate. The stimulus statements are given in Table 2.

Each session ended with a debriefing interview to investigate:

- participant’s views on the use of synthetic personae in e-retail applications
- their effective deployment on screen
- the characteristics required by such personae
- the conversational possibilities with personae in future applications.

Following the interview participants (1) ranked in order of merit their preferred top three sales assistants and (2) rated each sales assistant on a scale of 1 to 10 (ten being the best). At a later time seven participants attended a focus group which examined in greater depth participants’ perceptions and opinions about the synthetic personae, their effectiveness and how they might be deployed in applications.

| Questionnaire Items   | T1 | T2 | T3 | T4 | T5 |
|---|----|----|----|----|----|
| 1. I think this service is a good idea  | *  | *  | *  | *  | *  |
| 2. I think this service would be difficult for me to use.                             | *  | *  | *  | *  | *  |
| 3. I would like to use this service myself  | *  | *  | *  | *  | *  |
| 4. I felt the assistant was friendly  | *  | *  | *  | *  | *  |
| 5. I felt the assistant seemed competent  | *  | *  | *  | *  | *  |
| 6. The assistant's voice was not clear enough.  | *  | *  | *  | *  | *  |
| 7. I liked the assistant's voice.   | *  | *  | *  | *  | *  |
| 8. I felt the conversation was unnatural.   | *  | *  | *  | *  | *  |
| 9. I liked the appearance of the assistant.   | *  | *  | *  | *  |    |
| 10. I thought the assistant looked natural.   | *  | *  | *  | *  |    |
| 11. I thought being able to see the assistant was helpful.                            | *  | *  | *  | *  |    |
| 12. The appearance of the assistant was unsuitable for the home furnishings scenario. | *  | *  | *  | *  |    |
| 13. I looked at the assistant more than the living room.                              | *  | *  | *  | *  |    |
| 14. I felt the speech sometimes didn't match the lips.                                | *  | *  | *  |    |    |
| 15. I noticed the lips moving.  | *  | *  | *  |    |    |

Table 2: Questionnaire Items for Each Technology

### 3. RESULTS

#### 3.1 Overall Ratings

In order to obtain an overall rating of the ten synthetic personae, the results of the ratings provided during the post-experiment interview were analysed. The mean scores for each persona are shown in Figure 4. The videos were rated the best; the 3D talking heads were rated the worst.

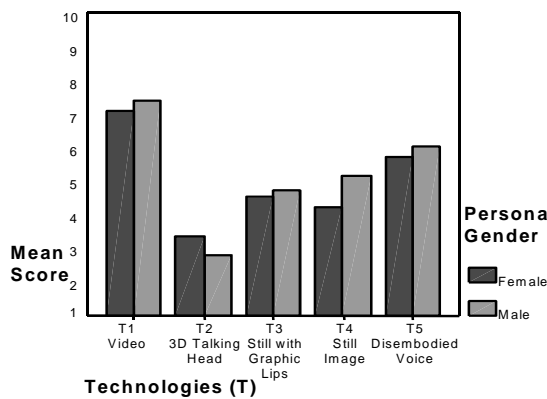


Figure 4: Ratings by Persona Technology and Gender

A 2 x 5 repeated measures ANOVA taking persona gender and persona technology as the independent factors was carried out. Standard multivariate tests such as Pillai's Trace showed highly significant differences due to persona technology ( $F = 21.88$ ,  $df = 4.0$ ,  $p < 0.01$ ) and significant differences for persona gender ( $F = 5.10$ ,  $df = 1.0$ ,  $p < 0.05$ ). There were no significant interactions between persona gender and technology.

Between-technology pair-wise comparisons showed similar though not identical patterns for male and female personae. With

respect to the male personae, the video (T1) and the disembodied voice (T5) were rated similarly and both were rated higher than the other three technology types ( $p < 0.05$  in all cases). The still image (T3) and still image with lip animation (T4) were rated similarly, and both were rated significantly higher than the 3D talking head (T2), both at  $p < 0.01$ . Participants' responses to the 3D talking head were poor. This technology was rated the worst of the persona technologies used in this experiment, ( $p < 0.01$ ). The results for the female personae were similar except that the 3D talking head (T2) was rated statistically the same as the still image (T4,  $p = 0.062$ ).

Due to the number of pair-wise tests carried out for this analysis, significance values close to 0.05 need be treated with caution. Consequently, the differences reported above between male T4 and T5 ( $p = 0.044$ ) and between female T3 and T5 ( $p = 0.025$ ) may not in fact be significant. However, the overall pattern remains unaffected with technologies T1 and T5 being the most preferred and T2 being the least preferred.

Analysis of the effects of between-subject variables showed there was no gender bias within the participant sample with respect to persona gender (i.e. male and female participants showed the same preferences for male and female personae). Similarly there was no participant gender bias towards persona technology. No effects were found for participant age group or for experiment supervisor.

#### 3.2 Analysis of Attitude Questionnaire Items

During the experiment an attitude questionnaire was completed, the individual items of which are discussed and analysed below. The results re-enforce the overall ratings presented above, but differ in interesting ways. It should be noted that Likert scales and ratings scales differ in the type of data they provide. The ten-point rating scale used was possibly more sensitive than the 7-point Likert questionnaire because participants were explicitly drawing comparisons between personae, rather than focusing their attention on individual personae when completing the Likert questionnaires. This may be the reason why certain items (i.e. Item's 4 and 5) produced results that differed from the trend suggested by the ten-point rating scale.

Three items in the questionnaire related to the users' attitudes to the service as a whole. The service was considered to be a good idea (mean = 5.42), and easy to use (mean = 5.52). Participants also agreed that they would use the service if it were available (mean = 5.05). Persona technology or gender did not influence these positive attitudes.

More surprisingly, the perceived friendliness of the assistant (mean = 4.53) and the perception of the assistant's competence (mean = 5.35), both of which were very positive, were also uninfluenced by either persona gender or technology. A tentative suggestion to the lack of variation between technologies here could have been due to the passive viewing nature of the experiment. Perhaps if the participants had interacted with the sales assistant, they may have had more informed opinions about friendliness and competence.

##### 3.2.1 Attitude to the Voice

With respect to the voice used in the service (one male and one female voice used throughout), the mean for item 6 indicated that participants found the voices of all personae were clear (mean = 5.72). There were no significant differences between persona types. Item 7 produced a mixed response as shown in Figure 5. A

2 x 5 ANOVA did not show significant differences due to persona gender ( $F = 2.82, df = 1.0, p = 0.104$ ) but did show differences for technology ( $F = 5.62, df = 4.0, p < 0.01$ ).

There was one significant result between the female technologies, that the voice of T1 was significantly preferred to T4,  $p < 0.01$ . The voice of the male video (T1) was significantly preferred to the male T2 ( $p < 0.01$ ) and T3, ( $p < 0.05$ ). Moreover, it was preferred to all five female personae, all at  $p < 0.01$ . Attitudes to the voice of male technologies T1, T3, T4 and T5 were significantly better than the voice of the male 3D talking head (T2), all comparisons at  $p < 0.01$ .

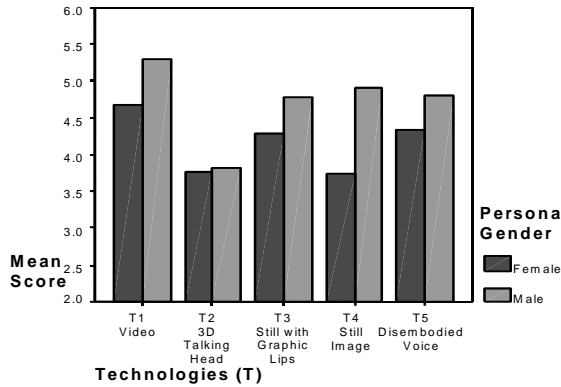


Figure 5: Attitude to Persona Voice (Item 7)

### 3.2.2 Naturalness of the Conversation

A 2 x 5 ANOVA for item 8 showed significant differences between technologies ( $F = 2.99, df = 4.0, p < 0.05$ ), but no significance differences for persona gender ( $F = 1.92, df = 1.0, p = 0.179$ ). Figure 6 shows that participants felt the conversation with the female 3D head (T2) was more unnatural than the conversation with the T1, T3 and T5,  $p < 0.05$ . This persona was also significantly lower than male technologies, T1, T4 and T5 (all at  $p < 0.05$ ). With respect to the conversation with the male personae, the still with graphic lip movement (T3) was considered to be the least natural with a significantly lower score than either the video (T1) or the disembodied voice (T5), both at  $p < 0.05$ .

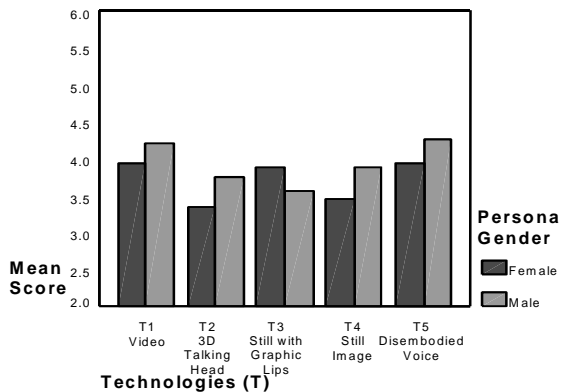


Figure 6: Attitude to the Naturalness of the Conversation (Item 8)

### 3.2.3 Appearance of the Personae

Five questionnaire items were included for use with technologies T1-T4. A 2 x 4 ANOVA showed that, with respect to the

appearance of the personae (questionnaire item 9) there were significant differences due to technology ( $F = 9.474, df = 3.0, p < 0.01$ ) and gender ( $F = 5.49, df = 1.0, p < 0.05$ ).

T-Tests showed the appearance of the male 3D talking head (T2) was significantly worse than T1, T3 and T4, all at  $p < 0.01$ . Similarly the female 3D talking head (T2) was significantly worse than T1 and T3, both at  $p < 0.05$ . The male video (T1) was significantly better than all other male technologies and in fact this persona was significantly better than all female technologies, (all at  $p < 0.01$ ). Both T1 genders were significantly better than both T2 genders, all four comparisons at  $p < 0.01$ .

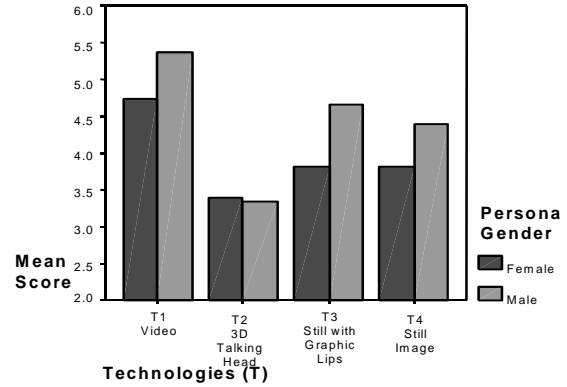


Figure 7: Attitude to the Appearance of the Assistant (Item 9)

Results from a 2 x 4 ANOVA indicated that for item 10 ('I thought the assistant looked natural') there were no significant results for personae gender ( $F = 0.029, df = 1.0, p = 0.867$ ), but there were significant results for persona technologies ( $F = 10.574, df = 3.0, p < 0.01$ ). On completion of paired sample T-Tests the following significant results were obtained. The male video (T1) was significantly favoured over T2 and T3, both comparisons at  $p < 0.01$ . The female video (T1) was significantly favoured over T2, T3 and T4, all comparisons at  $p < 0.01$ . In addition for both male and female personae, T3 was favoured over T2 (both at  $p < 0.01$ ). And finally the male still image (T4) was favoured over the male talking head, T2 ( $p < 0.01$ ).

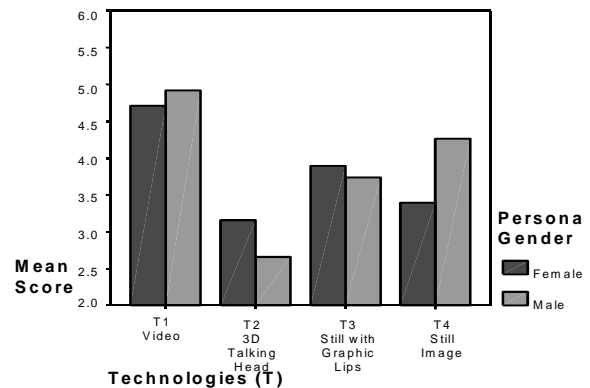


Figure 8: Attitude to the Visual Naturalness of the Assistant (Item 10)

Item 12 ("The appearance of the assistant was unsuitable for the home furnishings scenario") produced significant results with respect to persona gender ( $F = 3.04, df = 1.0, p < 0.05$ ), but no significant differences between persona technologies ( $F = 1.6, df$

= 3.0,  $p = 0.2$ ). In particular male technologies T2, T3 and T4 were significantly more suitable for the home furnishings scenario than the equivalent female technologies.

### 3.2.4 Helpfulness of the Personae

A 2 x 4 ANOVA of item 11 showed significant differences between technology ( $F = 2.55$ ,  $df = 3.0$ ,  $p < 0.05$ ). There were no significant differences due to persona gender ( $F = 1.26$ ,  $df = 1.0$ ,  $p < 0.272$ ). Participants felt that being able to see either the male or female videos (T1) was more helpful than being able to see either the male or female 3D talking heads (T2), all four comparisons at  $p < 0.01$ . Moreover for the male personae, participants felt that seeing T3 and T4 was more helpful than the 3D talking head, T2 ( $p < 0.01$ ). This is again comparable with the results obtained from the participant ratings in the post-experiment interviews, where the 3D talking heads (T2) were least liked.

### 3.2.5 Attention given to the Personae

Item 13 ('I looked at the assistant more than the living room') was asked with regard to the personae that were visible in the interface (T1, T2, T3, and T4). A 2 x 4 ANOVA showed that there was no overall significance between persona gender ( $F = 1.46$ ,  $df = 1.0$ ,  $p = 0.237$ ). There was significance between persona technologies ( $F = 3.14$ ,  $df = 3.0$ ,  $p < 0.05$ ), that is participants looked at certain persona more than the living room.

For the female personae, when the video (T1) appeared as the assistant, participants looked at the persona more than the living room and significantly more so than the 3D female talking head (T2,  $p < 0.05$ ) and the still image (T4,  $p < 0.01$ ). For the male personae significant results showed that participants looked less at the still image (T4) than the living room compared with other male technologies T1, T2 and T3 ( $p < 0.05$ ).

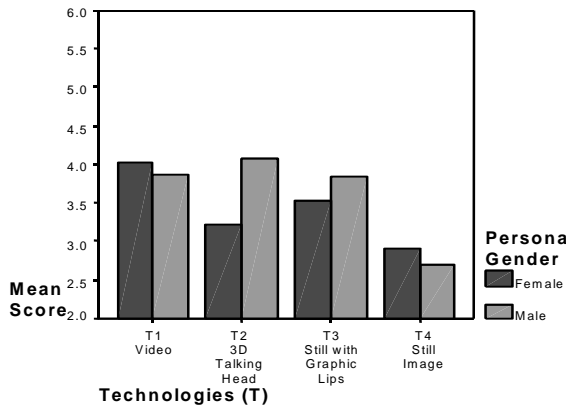


Figure 9: Visual Attention Given to each Personae (Item 13)

### 3.2.6 Lip Synchronisation

Two questionnaire items were related to the lip movement of the synthetic personae. Only three personae technologies had lip movement and hence a 2 x 3 ANOVA was used to analyse these items.

The ANOVA results for item 14 ('I felt the speech sometimes didn't match the lips') showed a significant difference for technology ( $F = 5.32$ ,  $df = 2.0$ ,  $p < 0.05$ ), but not for persona gender ( $F = 0.023$ ,  $df = 1.0$ ,  $p = 0.729$ ). Pair-wise comparisons showed that participants felt both male and female videos (T1)

had significantly better synchronised lip movement than the 3D talking heads, T2, and still image with graphic lip movement, T3 (both  $p < 0.01$ ). Moreover, the means for T2 and T3 were below neutral and therefore participants did not think the lip movement matched that well.

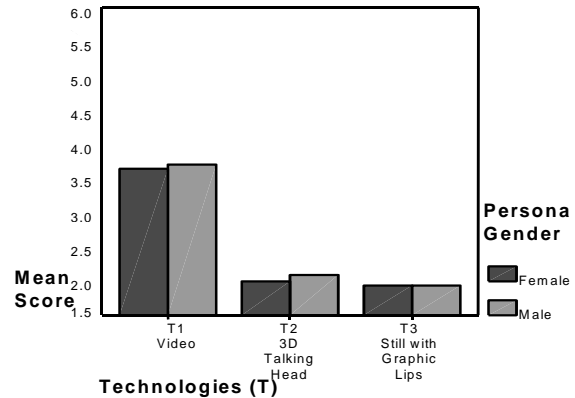


Figure 10: Lip Synchronisation (Item 14)

The results from item 15 ('I noticed the lips moving') showed that participants did notice the lip movement of the personae. A 2 x 3 ANOVA indicated that there were no significant differences within technologies ( $F = 0.804$ ,  $df = 2.0$ ,  $p = 0.46$ ) or between technologies ( $F = 0.94$ ,  $df = 1.0$ ,  $p < 0.34$ ).

## 3.3 Ranking and Rating

In the semi-structured interviews participants were asked to rank the best three assistants. The results are shown here in Table 3.

The male video received the majority of the votes from the 32 participants with the disembodied voices receiving over one third of the votes. There is an interesting relationship between the first and second preferences for the male and female videos (T1) and male and female disembodied voices (T5).

| Rank | No. of Votes T1 |    | No. of Votes T2 |   | No. of Votes T3 |   | No. of Votes T4 |   | No. of Votes T5 |   |
|------|-----------------|----|-----------------|---|-----------------|---|-----------------|---|-----------------|---|
|      | F               | M  | F               | M | F               | M | F               | M | F               | M |
| 1    | 6               | 13 | 0               | 1 | 0               | 1 | 0               | 0 | 6               | 5 |
| 2    | 11              | 5  | 0               | 0 | 1               | 2 | 1               | 3 | 3               | 6 |
| 3    | 5               | 5  | 1               | 0 | 4               | 5 | 2               | 6 | 0               | 3 |

Table 3: Preferences for Each Synthetic Persona

13 participants voted for the male video as a first preference. From this group of participants the majority of second preferences votes (8 votes in total) were for the female video. 6 participants voted for the female video first. 4 participants in this group voted for the male video second.

Similarly, of the 6 participants who voted for the female disembodied voice, a total of 4 participants went on to vote the male disembodied voice as their second preference. Finally three of the 5 participants who voted the male disembodied voice as their first preference, voted for the female disembodied voice second. In all cases over half of the second preferences votes went to the opposite gender of the same technology. Other first preference votes went to the male talking head T2 (1 vote) and the male still image with graphic lip movement T3 (1 vote). The highest amount of third preference votes (6 votes in total) were for

the male still image with graphic lip-synchronisation. The rest of the votes were distributed between the other synthetic personae. One participant failed to make a third preference vote.

### 3.4 Interview Feedback

Feedback from the semi-structured interviews indicated that there was a negative attitude towards the lack of facial expression. They thought the assistants should smile more to exercise more 'civility'. One participant wrote, "*the assistant's appearance looked very unnatural because of minimal facial expressions*". This suggests that usable applications need to achieve an appropriate level of facial expression.

People noticed that the lip synchronisation of T2 and T3 did not match the speech. Many thought this was distracting and annoying; these participants favoured either the disembodied voices, where they did not have to look at the head, or they preferred the video where the lip movement was more accurate.

There were a substantial number of negative comments about the female voice, but none for the male voice. For example the female voice was described as: "disinterested", "horrible", "tone was over confident", "her voice was harsh and irritating", "tinny". These comments raise interesting questions for the design of synthetic personae. A neutral, accent free voice may be more acceptable than the female voice used in this experiment.

All the participants who expressed preference for the disembodied voices (T5) commented that the picture of the sales assistant was distracting: having to look at the picture distracted them from the visual changes that were happening in the virtual environment. It is important to note that no matter how sophisticated the technology of the visual appearance of the assistant, if the interface demands too much attention from the user, synthetic personae may not actually enhance the service at all.

During the interview participants were asked to indicate the worst assistant and give reasons for their choice. 87.5% of votes were for the 3D talking heads; with 11 votes for the female talking head, 11 votes for the male talking head and 6 votes for both male and female talking heads. Obviously, the technological sophistication of these synthetic personae is far from acceptable with users indicated a marked preference to interact with more natural looking characters.

In summary, the most popular assistants were those which were most realistic and human-like viz. the male and female videos and the disembodied voices.

### 3.5 Focus Group Feedback

The first issue addressed in the focus group concerned opinions about the sales assistants. The focus group members made comments that the dialogue between the customer and the assistant didn't 'gel together' and seemed 'contrived'. Some participants felt that the assistant was impolite and 'pressurised the buyer'. To improve this aspect of the service, it was suggested that the assistants should have smiled more. In general focus group members felt that the assistant's appearance and voice, lacked emotion.

The focus group members were then shown short excerpts of the ten recordings and were asked to comment on each one. Comments made were that the female video, T1, had an unusual voice and it was apparent that she was reading from a script. There were no comments about the male video. Participants commented that the talking heads, T2, were 'awful', the female

head looked distorted and 'dummy' like. They said the male talking head had annoying head movements and seemed 'suspicious'. The female talking head 'distracted attention' from the home furnishings application.

Thirdly, comments about the still frame with hand crafted lip synchronisation, T3, suggested that the male assistant was not interested and was even seen as hostile. The female version was 'comical' and 'too unnatural'. The still image, T4, for both the male and female heads was distracting. Participants said that there didn't seem to be any point in having a still image. They expected the face to move.

Finally the disembodied voices T5 prompted comments that the female assistant appeared to be 'more confident' than the other female assistants. This is interesting considering the dialogue was identical in each video recording. Some members of the group would have preferred to see a face to match the voice. People who favoured the disembodied voice over other technologies suggested that the user should have the facility to remove the picture of the assistant after a certain time.

The participants were asked to suggest ways to improve the assistants. The issue of adding emotion to the face arose. In addition it was pointed out that hand gestures may enhance the performance of the synthetic personae as sales assistants. There was general agreement when participants were asked if they would like to see the sales assistant in the electronic retail environment as a 3D character. Interacting with cartoon characters was suggested.

## 4. DISCUSSION

The key findings of the experiment showed that:

- There was significant preference for the videos and disembodied voices
- People had a preference to interact with synthetic personae that exhibited facial gestures and emotion
- The 3D talking heads were rated the worst
- The male voice was preferred to the female voice

The aim of the experiment was to evaluate user acceptability toward a variety of male and female synthetic personae. The results showed that participants favoured the male and female videos and disembodied voices to all other technologies used in the experiment. It must be pointed out that it was difficult to gather more detailed information from the participants about the personae, as they did not actually interact with the sales assistant themselves; they merely overheard a recorded conversation.

The popularity of the disembodied voice raises interesting issues about the need for synthetic characters in interfaces. It seems that if the task is visually demanding the user may find a picture of the assistant distracting. It is therefore necessary for interface designers to assess services and applications carefully to establish if a synthetic persona is an actual enhancement.

Questionnaire item 7 retrieved information about user attitudes toward the synthetic personae voices. The same male and female voices were used for each set of five technologies however significant results showed preferences for the voices of the videos (T1) and disembodied voices (T5) suggesting the visual appearance of the persona has an impact on user attitudes toward the voice. This highlights an important cross-modal effect. The voices of T1 were most preferred, suggesting that the more natural

the facial movements, the more acceptable the voice. T5 was also significant here, raising issues also mentioned in the interviews, that if human-like personae do not exhibit natural facial expressions, participants may find them visually distracting.

Once it has been established that there is a need for a synthetic persona in an interface the next question is which type of persona? It is concluded from this experiment that people prefer to interact with more natural looking personae that exhibit human-like facial expressions, facial gestures and emotions. Essentially, the user demands that the system support natural face-to-face dialogue for it to be successful. The low ratings of the 3D talking heads suggest that the technology used to create them is underdeveloped.

During the focus group and semi-structured interviews participants were invited to make suggestions for improving the appearance of the sales assistants. Signals of friendliness (such as smiling) were given high priority for enhancing retail services. In general, facial expressions play a crucial role in attitude and perception of services. Lip synchronisation was highlighted: bad lip-synchronisation was found to be distracting and synthetic personae in which this was the case were undesirable. The choice of a suitable voice is also an issue that needs to be carefully considered.

The dialogues used for e-retail applications need to be examined more closely. The responses of the sales assistant impact significantly on the attitude responses of the participants. The output dialogue will contain a multitude of information about the personality of the persona. These are issues that will be even more important when creating fully functional systems. Virtual sales assistants must have traits that are conducive to selling and advising; stable, relaxed, sociable, conscientious, cheerful, patient and diplomatic.

This experiment was the first in a series. The next stage is to evaluate the same ten synthetic personae in a different e-retail environment. In this way the effectiveness of these synthetic personae will be investigated in a contrasting environment and it will be established if participants evaluate the same personae differently in this new environment. Secondly, the cast of synthetic personae will be extended to include five additional characters, which will include 2D and 3D humanoid cartoon-like characters. Based on issues that arose while evaluating the results of this experiment, a variety of voices or the voice of a professional actor may be used.

## 5. ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support for

this research from BT under its Strategic University Research Initiative, and the helpful planning of the experiment made by the research staff at BT Adastral Park. The work has benefited greatly from helpful discussions with colleagues at the University of Edinburgh, in particular Dr. John Foster.

## 6. REFERENCES

- [1] Andre, E et al., 'Personalising the User Interface: Projects on Life-like Characters at DFKI'. In Proc. Workshop on Conversational Characters, Tahoe City, California, 1998.
- [2] Biermann, A. et al. (National Research Council): 'More Than Screen Deep: Toward Every-Citizen Interfaces to the Nation's Information Infrastructure'. National Academy Press. 1997.
- [3] Bernsen, N. O.: 'Towards a Tool for Predicting Speech Functionality'. In Speech Communication, vol. 23, 1996.
- [4] Cassell, et al. J.: 'An Architecture for Embodied Conversational Characters'. In Proc. Workshop on Conversational Characters, Tahoe City, California, 21-30, 1998.
- [5] Foster, et al: 1998, 'An experimental evaluation of preferences for data entry method in automated telephone services'. In Behaviour and Information Technology, vol. 17, no.2, 82-92.
- [6] King, J. and Ohya, J.: 'The Representation of Agents: Anthropomorphism, Agency and Intelligence'. In Proc. CHI 1996.
- [7] Likert, R: 'Some Applications of Behavioural Research'. Paris, 1932.
- [8] Massaro, D: 'Perceiving Talking Faces: From Speech Perception to a Behavioural Principle'. MIT Press, 1998.
- [9] Maybury, M. T.: 'Toward Co-operative Multimedia Interaction'. In Lecture Notes in Artificial Intelligence 1374, Multi-modal Human-Computer Communication, 1998.
- [10] Parke, F.I. and Waters, K.: 'Computer Facial Animation'. A.K. Peters, Wellesley, 1996.
- [11] Shneiderman, B.: 'Looking for the Bright Side of User Interface Agents'. University of Maryland, 1994.
- [12] Thalmann, et al: 'Digital Actors for Interactive Television'. MIRALab, University of Geneva, 1998.
- [13] Thorisson, K: 1996, 'Communicative Humanoids: Model of Psychosocial Dialogue Skills'. Thesis MIT Media Laboratory, 1996.
- [14] Wyard, P et al: 'The MUESLI Multimodal 3D Retail System', ESCA Workshop on Interactive Dialogue Systems, 1999.

