Utilising Human Audio Visual Response for Lip Synchronisation in Virtual Environments

Liam Parker and M.A. Jack

Centre for Communication Interface Research
Dept. of Electrical Engineering
University of Edinburgh
80 South Bridge
Edinburgh
EH1 1HN

Tel: *0131 650 2771*
Fax: *0131 650 2784*
e-mail: *liam@ccir.ed.ac.uk*

This paper discusses the difficulties in maintaining the impression of a 3D virtual persona within a multimedia desktop environment and proposes a method in which the task can be minimised in such a way that it takes advantage of the level of human perceptual acceptance for asynchronous audio visual speech information. An example of the algorithm devised is tested both objectively and by subjective user appraisal and is shown to demonstrate promising results that merit further research.

Keywords: Lip Synchronisation, Speech

1. Introduction

With the increasing utilisation of novel multimedia interfaces through the use of user centric GUIs and desktop virtual reality it is important to examine the way in which users interact with new technologies within adaptable interfaces. One emergent technology gaining popularity is the use of virtual personae; computer generated anthropomorphic representations utilising lip synchronised synthesised speech. The advantage to fully synthesised character representations is that they are unbound from the user perspective; interactive dialogues can be fully directed by the user. Much work has yet to be done on how best to incorporate this type of technology into multimedia interfaces since they have generally required considerable computation power in order to maintain the impression of a virtual persona. Couple this with the need to guarantee a particular screen frame rate in order to maintain the synchronisation of the lips with the speech and it is evident that the demands such technology makes upon user's hardware is considerable. Since computation power and therefore rendered frame rates cannot be guaranteed within a multitasking multimedia environment these persona technologies are not necessarily suited to today's media rich applications. Also, if wide spread user trials are required for human computer interaction studies then the hardware cost in providing machines capable of sustaining a virtual persona interface in a multimedia application can become prohibitive. The real difficulty therefore in undertaking research in this area is the novel nature of this technology. There is a real need to produce an approach that allows persona interactions to be investigated within multimedia applications on technology that is not prohibitively expensive with respect to today's current user installed base.

2. Background

There are many psychological and user interaction reasons for using a computer synthesised face (Benoit, 1990) and extensive work has been done to construct both parameterised models that use minimum sets of variables (Parke, 1982) as well as extremely concise facial articulation models (Lee et al, 1995) that take account of the skull, muscles and skin. Using such synthesised faces together with text to speech and speech recognition systems for an interface is not new. In fact technologies where non-verbal and verbal communication coexist in a facial multimodal interface have been the subject of much research. A significant result of this has been the implementation of phoneme based talking personae (Waters, 1993 : Katunobu, 1995 : Ball, 1995) (computer generated graphical representations of humans) that can interact with a user.

Synthetic persona animation methods, such as DECface, require accurate synchronisation of the animation with the audio. With regard to the difficulties of accurate timings within multitasking multimedia applications it is interesting to examine animation approaches that take advantage of any perceptual bias user's have when they interact with talking personae in order to overcome synchronisation difficulties between the audio and visual aspects of virtual persona.

3. Modelling Issues

In order to produce a virtual persona capable of working as part of a multimedia application on desktop hardware a minimised facial animation system was devised. It was decided to design the animation system on a time based linear interpolated frame approach which is similar in function to the VRML 2.0 animation extensions (Pesce, 1995) since such an approach would allow any resulting persona algorithm to be utilised in World Wide Web applications. These animation extensions to VRML 1.0 are based on event specified time dependent motion and have been made extensible in order to allow other languages such as

Java (Flanagan, 1996) to interact with 3D modelled environments thereby allowing any algorithm to be readily portable to VRML applications.

Due to the demands placed on processing power by complex 3D environments and the considerable context switching that can occur in multimedia interactions complex algorithms such as Lee, Terzopoulos and Waters' (1995) were avoided. Even reduced parameter methods such as Parke's (1982) require a level of numeric calculation that is costly in real time on low specification desktop machines and difficult to adapt to constantly adjusting computation resource. Therefore, simpler methods of facial animation were considered. Also, in order to further optimise the system for desktop machines photorealistic texture mapping (Heckbert, 1986) was avoided and Gourad (1971) shaded polygons were used instead. Also, since the virtual persona is a generalised facial schema instead of an accurately modelled individual, the polygon count could be reduced while still giving the impression of a virtual person, thus further reducing computation power. Even with such a reduced quality rendering produced by a non-textured model it is interesting to examine the benefits of such a computer modelled face for the purposes of interacting with dialogue based tasks. The surface based facial schema used during system development and analysis is shown in Figure 1.

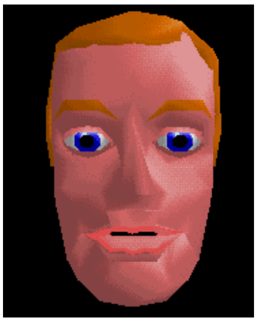
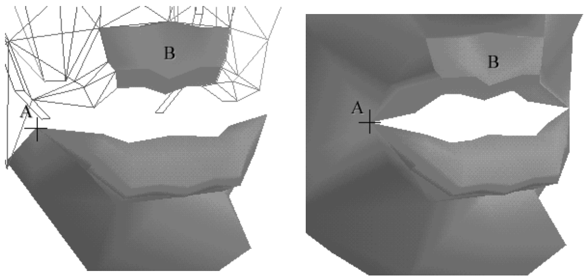


Figure 1 : The Virtual Persona



(i) (ii)

Figure 2 : Detail of jaw model

4. Problem Reduction

The central task in animating a computer modelled talking virtual persona is how to define facial states and the manner in which they are animated so that lip synchronisation can be achieved and emotional expressions can be modelled. Earlier work has used, among other things, vector sets, minimised data sets based on defined parameters (e.g. smile, blink, etc.) to control a subset of the model's vertices or a facial model based on the underlying anatomical structure of the skull, jaw and muscles. Using a full vector set to define precise states of a facial model would require a large amount of storage as well as reasonable computation power and is therefore unsuitable for present day network speeds (with regard to web based interactions) and user's hardware limitations. A minimised vector set based on specific parameters would use less storage, however the processing time required to move a large number of vertices for each frame restricts such a method from being used for a real time desktop system. A more reasonable approach is to build the modelled head through the use of many sub-objects based on the underlying anatomical structure of a human head. The main drawback to this approach is that in order to model slight changes in the face a large number of anatomical controls must be modelled. Also, a very complex skinning algorithm would have to be implemented in order to deal with such a complex set of sub-objects using up valuable processor power. Not only is this is a major consideration if the system is to be available for desktop machines but such control is not directly defined in VRML 2.0.

4.1 Reduced Complexity Facial Modelling

A design compromise was essential in order to allow the successful construction of a virtual persona using VRML based modelling. Any animation algorithm had to be linear time based in order for it to be as compatible with VRML 2.0 as possible as well as being simple enough not to disallow its use on desktop computers. The compromise taken was to employ linear parameters for controlling vertices where fine detail of movement is required and to employ anatomically based sub-objects for large deformation areas such as the jaw.

Figure 2 shows a section of the virtual persona jaw rendered with a smooth shaded algorithm (Gourad). In Figure 2i, point A is defined as a controlled vertex to allow fine adjustment of the corners of the mouth, for example when lip rounding is required. Point A is also defined as a 'skinning' point that exists in the skull and jaw sections. This can be seen in Figure 2ii where the jaw and skull sections of the modelled head are adjusted about vertex A. The upper lip section marked as B is defined as an anatomical sub-object since the majority of the upper lip can be approximated to moving as a single mass. The effect of this matching process can be seen in Figure 3 which shows the jaw open at varying degrees after the direct match 'skinning' has been applied.

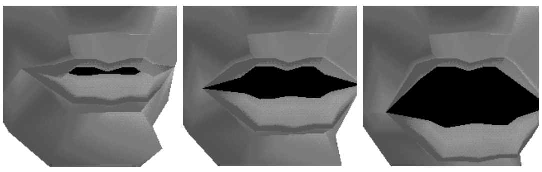


Figure 3 : Matched Vertex Jaw

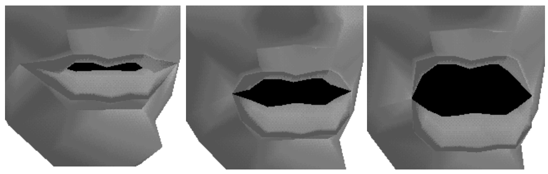


Figure 4 : With proportional adjustment

As can be seen from Figure 3 this direct matched 'skinning' has problems with complex entities like the lips. Therefore the matched vertices are adjusted with a linear relationship to sub-objects or other matched vertices in order to try to partially mimic the dependencies between various portions of the face. This is demonstrated in figure 4 where the skinning point *A* from Figure 2 has been adjusted inwards and upwards in proportion to the amount the jaw is opened.

Finally, the ability of the face to perform set actions is paramount for an interactive virtual persona. Actions require motions and motions require describing in the world model. A form of inverse kinematics could have been implemented but the deformations of the face required are relatively small and inertia can be discounted on this scale. Therefore, a simple linear interpolation approach was used whereby time based movements dependent on variables, a value or an object's property, can be defined. Each action is treated independently and multiple actions can be superimposed upon each other. For this initial application three lip state variables were defined relating to; the degree of openness of the jaw, the level of forward motion of the lips and, the degree of rounding of the mouth.

The animated lip synchronisation with a text to speech system output is achieved through the use of 'viseme' (Walther, 1982) mappings correlated to the output speech waveform. These visemes are defined as a set of actions, one for each phoneme in standard English, and are performed in the sequential order that the phonemes appear in the current utterance. Since the timing of the phonemes is known, each viseme action can be a simple functional linear motion that is timed to last as long as the current phoneme.

Through the use of visemes (visible phoneme mappings) the appearance of a computer head that is talking can be achieved. The main difficulties of timing visual actions with audible events are minimised by using a time based approach since the rendering is adjusted in real time so that visual events match with audible events producing a corresponding coarsening of the image's animation. This is less annoying to the user than a corresponding drop in voice quality or errors occurring with lip synchronisation.

4.2 Psychological Issues

The human psychological factors that have to be investigated within the field of computer generated interactive persona are many. It is not within the aims of this paper to address these issues. However, one particularly important aspect of visually cued speech perception is that of the effect of poorly synchronised visual and audio cues. It is not satisfactory within a multitasking, multimedia environment to expect a particular level of computation power for visual or audio channels or that the relationship between them will remain constant. Some degree of asynchronism will occur and therefore any animation algorithm devised should encompass the psychological confines involved.

One such psychological aspect with respect to virtual personae is the McGurk effect (1976). McGurk showed that there is a complex interdependency between visual and audio cues given to listeners when they observe a talking face. Massaro et al. (1996) showed that there is a discrepancy in the level of deterioration of intelligibility between leading and lagging visual information with respect to the audio speech. They show that the detrimental effect of asynchronous audio visual speech is greatest when the audio precedes the video rather than vice versa. When audio lags behind the video a delay of as much as 200 milliseconds remains unnoticed. However, when the video lags the audio a maximum delay of 100 milliseconds is acceptable. Therefore, there seems to be an acceptable lag window of 300 milliseconds within which any lip synchronisation animation algorithm should operate.

4.3 Overlapped linear visual animation

From a phonemically labelled database of 200 sentences of English speech the mean duration of the phonemes bounded by their standard deviation is shown in Figure 5. The maximum mean phoneme duration is for oi which is 180.03 ms with a standard deviation of 49.95 ms. This is within the 300ms lag window for audio visual lag. Therefore, the timing of precise closures and sub-phoneme duration movements of the lips can be disregarded since they should not greatly affect user’s perception of the system since they are of too short a duration.

Once sub-phoneme duration events are disregarded a purely linear interpolation between the viseme mouth states can be calculated. Since the algorithm is required to operate in a multimedia environment where frame rates will be variable due to varying application load no pre-computation of the interpolation is undertaken. At each frame the correct interpolated value of each variable is calculated immediately prior to rendering.

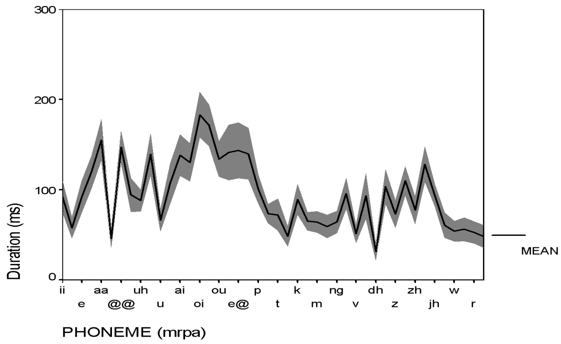


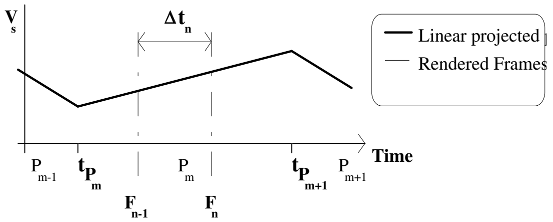
Figure 5 : English RP Phoneme durations (ms)

Through the use of a text to phoneme engine a phonetic stream is built along with the sequential time marker of each phoneme. The phoneme stream is then converted to an audio stream, through the use of a phoneme to speech engine, and is stored along with the original phoneme sequence. The audio and the phoneme streams make up the basis for an utterance that can be output immediately or stored for future playback. When an utterance is output the current time in milliseconds is stored (T_{begin}) and the audio data is played. The phoneme time markers are adjusted such that;

$$t_{Pm} = (T_{begin} + t_{Pm}) - 150ms$$

Where t_{Pm} is the time marker for the phoneme. This overlaps the visual time marker for the phoneme onto the previous audible phoneme. This is done in order that the timing used in the animation of the lips is close to the beginning of the audio visual lag window which begins when the visual aspect of the persona is approximately 200ms ahead of the audio. This gives a maximum allowable visual rendering lag of 250ms before the system would suffer from asynchronism outside the lag window.

During the audio playback, at each rendered frame (t_{Fn}), the interpolated value of each lip state variable (V_s) is calculated;



$$V_s = \begin{cases} t_{Fn} < t_{Pm} : V_{Sm-1} + \Delta V_{Sm}\left(\dfrac{\Delta t_n}{t_{Pm+1} - t_{Pm}}\right) \\ t_{Fn} \geq t_{Pm} : V_{Sm-1} + \Delta V_{Sm} \end{cases}$$

$[\bullet\ t_n = t_{Fn} - t_{Fn-1}]$ $[\bullet\ V_{sm} = \bullet\ V_{sm} - \bullet\ V_{sm-1}]$

Where V_{sm} is the value of the lip state variable V_s at phoneme *m*.

5. System Analysis

In order to analyse the benefits of the overlapped animation approach both objective data and subjective data had to be gathered. The system's audio visual lag had to be inspected with regard to the psychological confines. User's attitudes also had to be examined in order to assess whether the approximations in the animation were satisfactory.

The benchmark system used to test both the quantitative and qualitative aspects of the approach was a 90 Mhz Intel Pentium based personal computer with a standard (non-accelerated) PCI graphics card and 16 Mbytes of memory. This was considered by the author a standard desktop machine and not far removed from the type of machine currently being purchased by the average user. The polygon and vertex count for the Gourad shaded face model used during the analysis is shown in Table 1.

Part	No. of Instances	No. of Vertices	No. of Polygons
Skull section	1	237	257
Eye	2	180	161
Eyelid	4	52	38
Jaw	1	28	27
Top lip	1	24	20
Lower lip	1	31	28
Total		888	806

Table 1 : Face model statistics

5.1 Objective analysis

In order to investigate the effectiveness of the animation algorithm objectively frame rate and rendering load data were gathered. Frame rates for the virtual persona at resolutions ranging from 10000 pixels to 250000 pixels at 16 bit colour depth are shown in Figure 6.

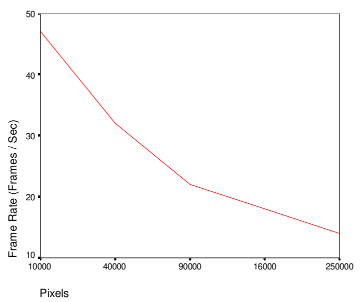


Figure 6: Frame rates for Virtual Persona

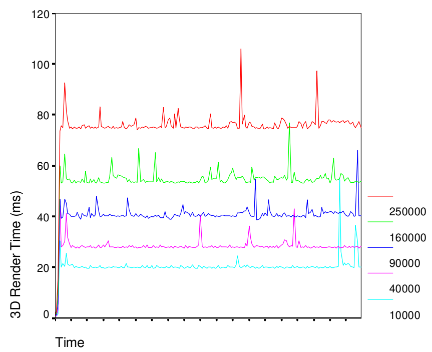


Figure 7 : Render times for Virtual Persona

At 10000 pixels the frame rate measured was 47fps. This is well in excess of film frame rates (24 fps) which are considered suitable for showing facial interactions. 22 fps are still achieved at 90000 pixels which is close to film frame rates. Taking the scenario where only one frame is rendered per phoneme we can calculate the maximum persona resolution allowed. The mean phoneme length in the speech database is 94.34 ms. This gives a minimum frame rate of 10.6 fps. This frame rate is still achieved by the virtual persona at 250000 pixel resolution. At this resolution the virtual persona would occupy half of an 800 by 600 resolution screen. Such a resolution range is considered acceptable by the author for multimedia applications since the persona system would have to simultaneously coexist with other on screen media.

It is not enough to consider the frame rate alone. In order to avoid the problems that asynchronous audio visual information can introduce the time taken to render the three dimensional facial representation after the viseme interpolation has been calculated must be less than the audio visual lag window. The time taken for virtual persona animation frames to be rendered at a range of resolutions is shown in Figure 7.

It can be seen that the render times are relatively stable at any particular resolution. The extreme peaks evident at every resolution level coincide with the beginning of a new phoneme and audio stream and represent the computation of the phoneme time markers and the operating system multitasking between the persona system and the audio stream. These peaks effectively occur between vocal prompts and therefore do not degrade animation during lip synchronisation itself. The average render time at 250000 pixel resolution is less than 80 ms which is well within the 300 ms lag window.

5.2 Subjective analysis

In order to test user attitudes towards the virtual persona system a precisely controlled usability experiment incorporating this system was carried out. It was decided to carry out the initial experimental benchmarking of the architecture in a laboratory set-up rather than a networked web based trial. Therefore a repeated measures design could be used since it would be difficult to constrain a fully networked trial under a repeated measures based experimental definition.

User attitudes had to be carefully considered. In order to avoid confounding any dependent variables in the experiment, in particular detrimenting attitudes towards the persona system through possible intonation inaccuracies in the text to speech system, real voice prompts were recorded and the initial phonetic timings were marked off-line by phonetic segmentation by hand. This produced a series of timed phonetic strings, one for each recorded prompt, which were fed to the virtual persona agent as the audio prompt was being played back.

Since such careful attention had been given to the voice, similar considerations were required for the speech recognition. This is particularly relevant when gathering empirical data on a subjective dependent variable such as general attitude. Poor recognition performance could degrade user attitudes to the extent where the dependent variable was confounded with variations in recognition and no meaningful data about the virtual persona system could be gathered. To avoid this problem a Wizard of Oz (WOZ) methodology (Jack, 1992) was used to simulate the speech recognisor with 100% accuracy.

In order to avoid type 1 (concluding sigificance where ther is none) errors when carrying out T tests (standard statistical significance tests) on subjective questionnaire data a mean response score was taken over a number of general statements about the interface.

5.2.1 Experiment Design

The experiment to investigate user attitudes towards the virtual persona agent was based around a simple shopping context with which the users were comfortable. The system prompts requested basic information from the user such as product cost and type. A relevant product description was then presented to the user which could then be accepted or not. In order to concentrate user attention on the synthetic persona, plain graphical text was the only other medium presented on screen. The text was used for encouraging the user to speak words from the recognition system vocabulary and participant priming consisted of an information sheet stating the context of a holiday booking teleshopping service.

To measure participant attitude towards the virtual persona a repeated measures group design was used with the independent variable (IV) being a binary state representing the presence or not of the persona representation. Order effects were adjusted for by presenting the IV conditions in one order to a random sample of half the participant population and in reversed order to the remainder. The interface without the persona present visually was simply a variation where the persona object was moved off screen. Therefore, the experimental conditions consisted of one condition which presented the persona visually and audibly plus text prompts, and another with just the text prompts plus the persona's voice. Each user experienced both conditions, with randomised order of presentation, and after each use, a usability questionnaire was completed. To fully define the operational conditions without constraining the participant to an unnatural form of behaviour, the experimental task was set to allow participant choice and browsing while also maximising exposure to the interface. The participants were asked to listen to three holidays of their choosing before accepting one. The priming was identical irrespective of the experiment condition being undertaken.

In order to gather objective response data the Wizard of Oz simulator system was set to record interruptions and user choices. For subjective attitude data a 7-point Likert (Oppenheim, 1966) response scale questionnaire was used. Through the use of pilot studies and literature review (Poulson, 1987 : ISO, 1990) a broad set of service attributes and user responses important for evaluating the usability of human computer interfaces were identified. These attributes include ease of use, perceived reliability and efficiency, perceived friendliness of the service, the degree of control over the service users felt they had during their interaction, and the degree of frustration experienced while using the service. These service attributes and user responses were selected to cover the largest number of usability dimensions relevant to dialogue based interfaces generally.

In addition to the general measured features, seven extra questions relating specifically to the virtual persona and the ability to interrupt were included. The questionnaire was designed as a randomised sequence of stimulus statements, one for each of 25 features, with associated 7-point Likert response scales. The Likert responses were scored from one (strongly disagree) to seven (strongly agree) with four being the neutral mark. Any negative Likert question's scores were inverted about the neutral mark (ie. scored from strongly agree to strongly disagree) in order to provide statistical coherence to the data collected. The participants completed a questionnaire after each experiment condition and the mean response across the 27 Likert questions was taken as a general attitude scoring for the dependent variable. This gave a quantitative value to the experimental effect allowing statistical analysis of the results.

In the experiment the WOZ server handled the delivery of introductory and interactive dialogues along with synchronised animation images of the talking persona and multimedia presentations. The software also registered keystrokes from the experimental operator made in response to spoken input from the participant and was set to respond with 100% recognition accuracy. All data on the experimental operator's inputs were stored and the required response prompts were output.

A participant population was selected for use in a number of experiments by a market research company in such a way as to reflect the make up of the general populace in age, gender and social grade defined by occupation (Social Grading Survey). The experiment participants were taken as a random sample of this recruited population.

5.2.2 Experiment Results

The overall mean values for the Likert questionnaire responses are shown for each of the four experimental cells in Figure 8 with their probability values. Taking the core Likert questions across the 'no head' and 'head' conditions in order to remove order effects an applied pair wise sample t-test results in $t = 1.232$ with 11 degrees of freedom (dof) and a probability (p) equal to 0.24.

Looking at the naive use, an independent pooled variance t-test results in $t = 0.667$, dof = 10 and $p = 0.52$. The second use cases have $t = 1.621$, dof = 10 and $p = 0.14$ for an independent pooled variance t-test.

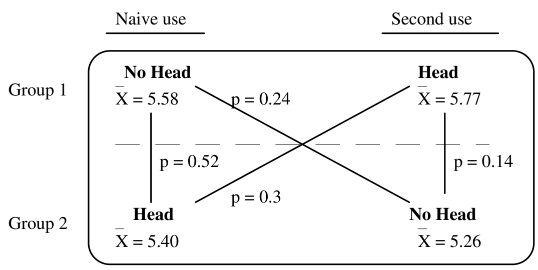


Figure 8 : Mean Likert responses

Noting that the attitude means of the 'no head' cells' reduce from the naive to the second use an independent pooled variance t-test results in $t = 1.257$, dof = 10 and $p = 0.24$. Looking at the increase in mean attitude scores from the 'head' cells an independent pooled variance t-test produces $t = 1.09$, dof = 10 and $p = 0.3$.

5.2.3 Subjective Experiment Discussion

The statistical data resulting from the subjective questionnaire responses has to be taken as an initial exploratory examination of user attitudes to the virtual persona due to the small number of subjects (12) used in the experiment. Therefore, a statistical significance in excess of 95% was unlikely due to any effects being tested for not being of such magnitude as to be readily observable with such a small participant population. However, significances just below this level should not be discounted but could be initial indicators that a network based field trial could be worth the effort.

In order to measure any user attitude influence the virtual persona had on the interface the Likert user response data was inspected. Comparing the ‘no head’ with the ‘head’ conditions gives an increased level of positive user attitude to the head version but it is not statistically significant. This may be due to any attitude experimental effect is small coupled with the relatively small size of the participant population. However, it is interesting to note that the attitude towards the system becomes more positive from the naive ‘head’ use to the second ‘head’ use even though the mean user attitude for the naive ‘head’ use is lower than the naive ‘no head’ use. Also, the ‘no head’ cells attitude’s fall from the naive to the second case. This is an interesting finding which will be tested in future experiments.

The mean attitude from the naive cells show that users were moderately positive towards both versions of the service however after the second use, after which they had experienced both ‘no head’ and persona versions, the mean attitude score decreased for the service with no persona present and increased for the persona version. It seems that on seeing the virtual persona interface users were moderately positive but it may be that once they had both experiences to compare they preferred the persona to the version with no persona. Though this would have to be confirmed during a full network trial.

6. Conclusions

It has been shown that a viseme animation algorithm can take utilise phoneme overlap to take advantage of an inherent audio visual lag window in order to optimally operate under asynchronous multimedia conditions. However, measures for the effectiveness of a particular animation algorithm are difficult to evaluate. Although testing of the algorithm at varying resolutions for both frame rates and render lag times shows that the system can operate suitably on desktop hardware while remaining within the lag window it is difficult to draw conclusions about the true effectiveness of the algorithm from this data alone. While the reduced complexity facial modelling used in this example is too impoverished to enable true lip reading it has been shown that initial subjective user trials have yielded encouraging results. These experiment results, although conducted on a small population, can be viewed as vindication for a full network trial in the future. In order for a future trial to fully assess the advantage of this phoneme overlapped animation it would have to determine whether user’s perception of verbal prompts is affected by the visual presence of the virtual persona. Such a trial is planned for the near future.

It is important to note that utilising human perceptual acceptance of asynchronous audio visual speech is not limited to low specification applications. This approach will aid lip synchronised speech perception in heavily loaded virtual environments on high specification hardware when undue loading produces periodic jitter in the animation frame rate.

Acknowledgements

The authors wish to acknowledge the help of Dr. John Foster of the University of Edinburgh and Dr. Fred Stentiford of BT Laboratories, Ipswich UK who have established the basic Wizard of Oz usability assessment methodology used in this work.

References

Ball, J. Eugene and Ling, Daniel T. - Spoken Language Processing in the Persona Conversational Assistant : ESCA Workshop on Spoken Dialogue Systems. 1995.

Benoit, Christian. - Why Synthesize Talking Faces? : Proc. of the ESCA workshop on Speech Synthesis, Autrans, France. 1990.

Flanagan, David - Java in a Nutshell : O'Reilly & Associates, Inc, 1996.

Heckbert, Paul S. - Survey of Texture Mapping : IEEE CG&A. 1986

Gourad, H.. - Continous shading of curved surfaces : IEEE trans. on Computers, C-20(6), June 1971, pp623-629.

International Standards Organisation, Ergonomic Requirements for Office Work and Visual Display Terminals (VDTs) : ISO CD 9241-11 (1990)

Jack, M.A. , Foster, J.C. , F.W.M. Stentiford - Intelligent dialogues in automated services. : Proc. International Conference on spoken language processing (ICSLP-92) pp715-718, 1992

Katunobu, Itoa , Hasegawa, Osumu, et al - An Active Multimodal Interaction System : ESCA Workshop of Spoken Dialogue Systems. 1995.

Lee, Yuencheng , Terzopoulos, Demetri and Waters, Keith - Realistic Modeling for Facial Animation : SigGraph. 1995.

Massaro, D.W. , Cohen, M.M. , Smeele, P.M.T. - Perception of Asynchronous and Conflicting Visual and Auditory Speech : Journal of the Acoustical Societyof America, Vol. 100, No. 3, pp1777-1786, 1996.

McGurk, H.- Nature, Vol. 264, pp746, 1976.

Oppenheim, A.N. - Questionnaire Design and Attitude Measurement : (Heinemann, London), 1966.

Parke, Frederic I. - Paremeterised Models for Facial Animation : IEEE CG&A, pp61-68. 1982.

Pesce, Mark - VRML browsing and building cyberspace : New Riders Publishing, 1995.

Poulson, D. Towards Simple Indices of the Percieved Quality of Software Interfaces : In IEE Colloquium - Evaluation Techniques for Interactive System Design. IEE, Savoy Place, London (1987)

Social Grading on the National Readership Survey : NRS. A JICNARS.

Walther, E.F. - Lipreading : Nelson-Hall Inc, Chicago, 1982.

Waters, Keith and Levergood, Thomas M. - DECface: An Automatic Lip-Synchronization Algorithm for Synthetic Faces : Digital Equipment Corporation Cambridge Research Labs Technical Report Series. CRL 93/4 - September 23, 1993