Technological University Dublin

ARROW@TU Dublin

Conference papers

Digital Media Centre

2009-01-01

Reusable, Interactive, Multilingual Online Avatars

Charlie Cullen
Technological University Dublin, charlie.cullen@tudublin.ie

Carl Goodman
Technological University Dublin, carl.goodman@peppersghost.com

Paula McGloin
Technological University Dublin, paula.mcgloin@tudublin.ie

Anna Deegan
Technological University Dublin, anna.deegan@tudublin.ie

Follow this and additional works at: https://arrow.tudublin.ie/dmccon

Part of the Other Computer Engineering Commons

Recommended Citation

REUSABLE, INTERACTIVE, MULTILINGUAL ONLINE AVATARS

Charlie Cullen[1], Carl Goodman[2], Paula McGloin[1], Anna Deegan[1], Evin McCarthy[1]

[1]Dublin Institute of Technology, Ireland, charlie.cullen@dmc.dit.ie

[2]Peppers Ghost Productions, England, carl.goodman@peppersghost.com

Abstract

This paper details a system for delivering reusable, interactive multilingual avatars in online children's games. The development of these avatars is based on the concept of an intelligent media object that can be repurposed across different productions. The system is both language and character independent, allowing content to be reused in a variety of contexts and locales. In the current implementation, the user is provided with an interactive animated robot character that can be dressed with a range of body parts chosen by the user in real-time. The robot character reacts to each selection of a new part in a different manner, relative to simple narrative constructs that define a number of scripted responses. Once configured, the robot character subsequently appears as a help avatar throughout the rest of the game. At time of writing, the system is currently in beta testing on the My Tiny Planets website to fully assess its effectiveness.

Keywords: avatars, interaction, multilingual, intelligent objects.

1 Introduction

The development and implementation of embodied conversational agents [1, 2] and avatars [3] is well documented in existing literature. Common approaches range from fully automated systems that generate all content programmatically through to semi-automated systems that leverage existing content such as input text or even pre-recorded speech in the delivery of the final avatar. Fully automated systems [4, 5] employ plan-based approaches to the generation of content for both single avatars and teams of presenters. Such systems create both the speech and gestures for a given avatar, often using automatically scripted synthesised speech content as the basis for gestural events in the final delivery.

In the case of semi-automated systems that employ pre-authored material, some form of manual or automatic annotation of that content is required to inform the rendering stage. For example, the BEAT system [6] performs a shallow natural language analysis to convert input text into synthesised speech. The system automatically annotates the text with gestures based on rules that are derived from studies of human dialogue, but does not provide the option of real speech input. Piwek and colleagues [7] automatically generate non-verbal behaviours from discourse analysis of text to create dialogues between embodied conversational agents. Their system finds rhetorical relations in an input text and maps them to question-answer pairs that are used to convey that information using avatars with synthesised speech.

Some systems do utilise real speech in the creation of avatars, though no implementation of the aforementioned gestural systems is present in existing work. LivingActor[TM] [8] can utilise both real or synthesised speech (from input text) to produce automatic lip-synching for animated characters, but the inclusion of gestures is not possible. The final character is output in a variety of video formats (AVI, MPEG, FLV) for distribution either online or using mobile devices, and in this regard indicates the great potential of avatars in online content delivery. Having said this, the prioritisation of movements and gestures in human communication [9] suggests that an avatar must perform much more than lip-synchronisation in order to communicate effectively with the user. The system proposed in this paper seeks to integrate the allocation of gestural events as employed by the BEAT system with the analysis of real speech for online distribution as with the LivingActor[TM] system. In so doing, the system aims for the production of avatars with higher quality real speech that possess a greater capability for gestural communication with the user.

2 Production Requirements

The inclusion of interactive multilingual avatars in the '*My Tiny Planets*' website [10] was initially suggested as a means of assessing the scope of character animation development undertaken as part of the Salero [11] project. In the project, the use of speech analysis to separate the animation authoring and rendering processes had led to the possibility of developing a system that would allow online interaction with an avatar to take place in any language within a pre-scripted domain. Further, the development of an online avatar player using Adobe Flash and Flex AS3 code had also suggested how the avatar could be configured live by the user at runtime, with the long-term aim of developing a multi-character system that would allow reusable assets to be integrated into production.

In developing such a system for online avatars, several core criteria were considered. The reuse of assets was considered crucial to an effective and scalable workflow, allowing speech and animation content to be leveraged in different contexts without extra production effort. In order for a useful avatar to be developed, means by which the avatar could

interact with the user were also required. To develop a truly scalable system for production, the ability of the avatar to function with input from multiple languages was also needed.

2.1 Reusable Assets

The reuse of speech and image assets was a core goal of the prototype development, allowing animation authoring and rendering to be split into 2 completely independent processes. Thus, if a particular speech asset has been authored and stored in the content database, it can then drive any avatar available to the system without further production work. This is considered to be an important innovation on current online animation processes, providing a method where scripted narratives form the core structure of animation authoring that can be performed quickly and cheaply. By separating the avatar and its movements from the sequence of movements assigned to a specific speech asset, the development of a character can thus be performed independent of the specific context it may eventually be employed in. This method aims to facilitate the create of intelligent objects for character animation [11] that possess personalities with associated behavioural traits that can be reused in a variety of productions and contexts. In so doing, this method aims to allow animators to spend more time on the detail of a specific movement or gesture, rather than having to divide their efforts in repeating each gesture every time it is required during a speech asset.

2.2 User Interaction

The capability of the avatar to interact with the user was also important. In the case of an online game site for children, the means by which the user can receive varied feedback was considered to be important to the perception of the avatar as 'real'. In particular, the inclusion of light hearted phrases (section 3) reacting to the selection of certain body parts was seen as a way of entertaining children using the robot, ideally introducing the concept of an avatar that will appear during the game to give feedback in a manner that will be welcomed by the user. The specification of this feedback was also constrained by the use of real speech assets- thus precluding a completely open domain for user interaction as in the case of speech synthesis. The decision to work with real speech was made largely on the quality of currently available open domain systems, alongside the use of specific acoustic analysis methods (section 5.1) that would not function well with concatenative speech synthesis. The decision to work with a limited domain gave no production advantage for real and synthesised speech, where the recording time for either narrative driven assets or a dedicated synthesis corpus was not significantly different. Interactions are defined relative to a scripted narrative designed to cover possible user operations in the prototype. The main interaction section of the system occurs during the avatar configuration phase, which was of particular interest for a children's website. By responding to the users choice of avatar body part it was intended to create an engaging environment for children to play with the avatar.

2.3 Multilingual Development

The creation of an avatar that would function independently of input language was also considered important. Language independence was achieved by employing a purely acoustic speech analysis method to drive the animation authoring process (section 5.1). Although no linguistic information is included automatically, descriptive metadata relating to each asset can be added as a transcription if required. Although not implemented at present, it is envisaged that a larger scale system would make extensive use of search and retrieval methods to build combinations of scripts and phrases from an initial production specific core. By controlling the animation using acoustic cues, language independent avatars can be implemented. In so doing, an asset in any given language can thus be authored by a speaker of that language and then used by any avatar available to the system during production. As different characters and languages are added to the system, no additional configuration or production work is required.

2.4 Prototype System Workflow

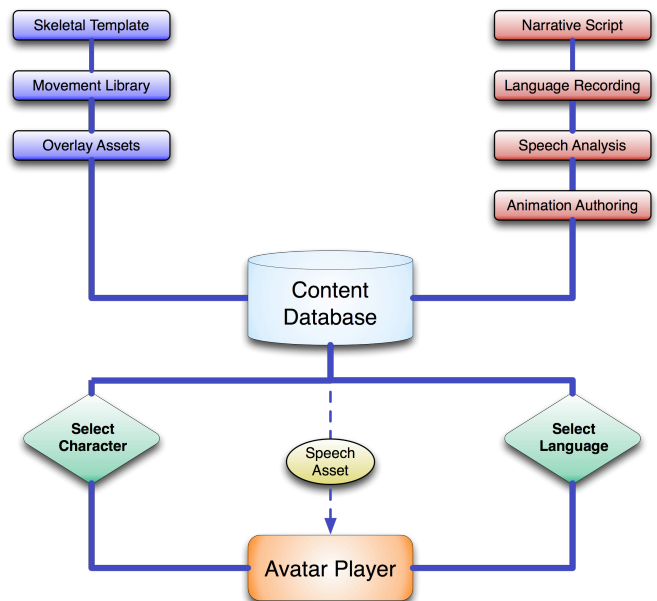The prototype system workflow is as follows (Figure 1):

Figure 1: Production workflow architecture for the avatar system. The diagram defines both the audio and animation content required to implement the avatar. A speech file in a given language is allocated movements and stored in the database. An avatar consisting of an IK skeleton with a library of gestures is stored alongside potential overlay assets. When the user requests a specific character in a certain language, the relevant speech asset is retrieved from the database and rendered by the player.

In the first section of the workflow an avatar comprising of a skeleton and its associated movements is created. This avatar is considered as an intelligent object [11] in terms of its physical capabilities, personality and behaviour. This base avatar can then have overlay assets associated with it that will form the final character, all of which are stored together in the content database. In the second section of the workflow, speech recordings (in any desired language) of narrative based scripts are analysed and then authored to append gestures and movements related to the events in the speech. This authored asset is then available for rendering using the avatar player and is also stored in the database. In the output stage, the player application can combine any available character with speech assets from any language for output to the user. The player can also implement scripted interactions (section 3) using a set of playback indices within a single file. For example, when the user selects a new head for the Robbie avatar in the prototype system a specific index to a head phrase is accessed by the player. This phrase is then played in accompaniment with the movements and gestures allocated to events in that portion of the speech asset during the authoring stage. This allows a limited domain of interactive phrases to be produced that can then be accessed at runtime and rendered through the player, allowing user selections to control the reactions and behaviour of the avatar.

3   Narrative Construction

In the case of an interactive online gaming environment for children, a simple set of narrative possibilities has been used to prototype the system. In the most basic operational workflow a user will select their chosen language, configure an interactive avatar using the available assets and then commence playing the game. From this point onwards, the role of the avatar is to provide information specific to aspects of the game during actual gameplay. With this in mind, 3 scripts were constructed:

- *Avatar configuration script*- this script is the most complex of the 3, as it plans the possible responses of the avatar to user selection. Thus, if the user selects a certain head for the avatar the avatar must provide a random response from the available phrases to the selection. By recording a discrete set of phrases for each body part (Table 1), a playback index can be used to select them at random during gameplay.
- *Game welcome script*- this script is used to introduce the user to the game level, informing them that the avatar can be found within the game to give them further instructions.
- *Instruction script*- at certain points in the game, the user will find their avatar situated in the game. By clicking in the avatar, pertinent information is then delivered by the avatar.

The main interactive element of this workflow is the avatar configuration script, containing responses to the user customising their avatar from the available body parts. A list of 10 phrases for each body part (head, arms, legs and torso) was compiled that could be delivered by the avatar at random in response to user selections (Table 1):

	Head	Torso	Arms	Legs
1	Is this the face you had in mind for me?	A new shiny body, straight from the factory.	These arms seem flexible yet fashionable.	What wonderful shoes, I like the colour.
2	What happened to my head?	This coat is just the thing for exploring new planets.	I shall be able to lift many things with these arms.	Do my knees look knobbly to you?
3	That was weird, I felt lightheaded for a second.	Yes, I think this colour suits me perfectly.	Are these ok- do you like the way they look?	I can dance!
4	Oh, I'm starting to feel a bit dizzy!	Oww, careful when you do that!	I think I can wave to you with these.	I wonder if I can jump in these shoes- wahey!

Table 1: Selection of configuration script phrases- *The table above give examples of some of the interaction phrases recorded for each body part. The aim of these phrases is to suggest to the user that the avatar is reacting directly to the choices they make, thus engaging them in an interactive configuration process.*

Although generic in the prototype stage, a more complex rule based system is planned [12] that links certain phrases to specific skeletons and assets to give a greater sense of interaction. By augmenting a general set of phrases that relate to all characters with more specific content linked to skeletal shapes (or even unique assets), the aim is to develop interaction scripts that are as flexible and varied as possible. For example, in a larger scale production there may be many different characters with differing torso shapes in many different colours. By linking a specific colour for every character, the avatar could directly reference that colour in response to a given torso asset. In contrast, a robot skeleton with wheels instead of legs would never reference knees (Table 1), so certain phrases would be prohibited for that character.

4   Avatar Development

In creating an interactive avatar for the My Tiny Planets website, discussions were held with the producers to ascertain what type of character would be best suited to user customisation and information delivery. It was decided that several existing robot characters would provide good scope for adaptation and customisation as different versions of the '*Robbie*' robot help avatar. One of the advantages of using a robot character was that children would be less expectant of photo-realistic movements and behaviour, thus allowing the

avatar to be more entertaining in its movements and gestures than could be achieved using a human figure. For example, while a robot can dance in a jerky manner and appear comical, a human avatar performing the same movements would perhaps be perceived by the user as clumsy. As the avatar was intended to convey information elsewhere during the game, it was important that it appeared as familiar and knowledgeable rather than awkward. With the prototype characters for the avatar defined, a skeleton for the avatar was built for animation.

4.1 Skeletal Templates

The initial set of 8 robots developed by Peppers Ghost Productions were redesigned to fit a specific skeletal shape and dimensions, thus allowing the user to load in different body parts at runtime. This process required the 3D models used for each robot character to be edited so that joint dimensions (e.g. shoulder width/leg length) were consistent in all cases. The Robbie avatar could then load assets from each of these robot characters interchangeably, thus providing a reusable workflow that could be scaled to include many other robot characters found in the 'My Tiny Planets' site. Although future development will consider a wider variety of characters and skeletal types than variations on robots for children's entertainment, the skeletal template used in the prototype conforms to a simple 2D biped skeleton that allows development of the avatar to be undertaken using common human gestures and movements (section 4.2). The avatar skeleton was built in Adobe Flash CS4, allowing the avatar's joints to be defined as inverse kinematic (IK) links using the CS4 bone tool (Figure 2):
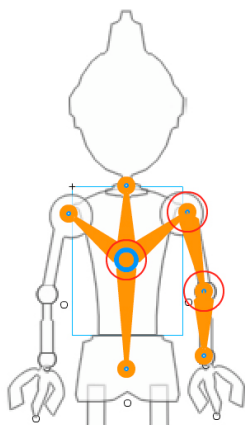
Figure 2: Joint definition using IK links in Adobe Flash CS4. The joints connect flat PNG image assets corresponding to each limb in the avatar, all of which can be swapped at runtime.

Each link in the chain is associated with a dedicated image asset, with rotation points being specified at joint intersections as illustrated. The blank assets used in the avatar construction stage are then swapped for specific assets taken from the robot templates by the user when configuring the character. The blank skeleton was used to create a set of

movements and gestures that can be triggered at runtime by the animation authoring script. In defining these animations, a frame label was used to demark the beginning keyframe of a specific movement for authoring purposes. When subsequently loaded into the authoring tool (section 5), these labels define the possible movements available for a given avatar. In this way, any skeleton created for use with the system can include its own unique library of movements and gestures specific to the capabilities of that skeleton (Figure 3):
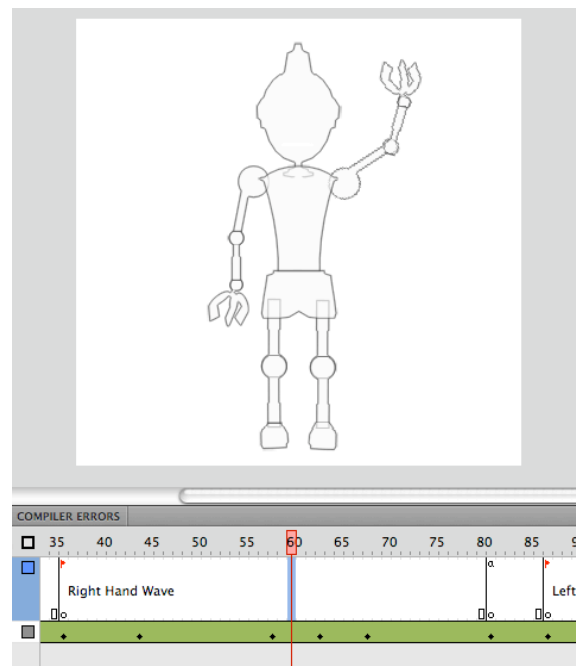
Figure 3: Frame label movement demarcation in AS3. Each movement begins from an initial position, allowing combinations of movements to be performed during an animation sequence. When a specific movement is defined in the authoring stage, the frame label is used as a trigger in the avatar player.

4.2 Movement Libraries

The creation of a reusable avatar focuses on the implementation of defined movement sequences in each case, so that commonalities can be exploited between different characters. For example, any avatar with arms and hands can perform a general waving gesture, while avatar head movements may vary in speed and distance but still retain common direction. Although the development of a comprehensive movement library is a complex (and ongoing) task, production discussions considered how best to develop and initial library based on the constraints of the Flash CS4 IK bone structure. Although the core inclusion of IK is a major step forward for the Flash development suite, it is still a long way from the subtlety and flexibility of dedicated 3D rendering engines. For this reason, a small set of simple and verbose gestures were developed that could be applied to specific events of importance within a speech asset, rather than progressing with the more subtle rhythms and beats [6]

possible with tools such as 3DSMax and Maya. As a result, the prototype avatars contain 20 movements that can be used during the authoring process (Table 2):
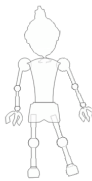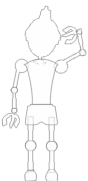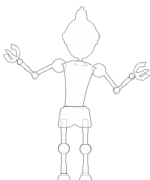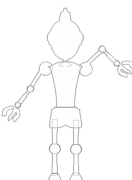
Thinking	Left Hand Wave
Hands on Hips	Right Hand Wave
Dance	Head Scratch
Two Hand Wave	Dance 2

Table 2: Example movements in the avatar prototype- *The movements shown are examples of those that can be performed by the avatar. Further extension of the movement library in the final system will allow a greater degree of subtlety and complexity to be employed during the authoring process.*

Although not available in the prototype, future work relating to groupings of movements and gestures will consider how to map movements from a certain type of avatar onto another avatar with different capabilities. For example, if a character with wheels is used then leg movements used by other avatars will have to be reinterpreted to suit the new avatar skeleton. In order to reuse an asset authored for one avatar with others of different skeletal abilities, semantic relationships (or perhaps an ontology of gestural concepts) between groups of gestures and movements will be required. Such a structure would ideally leverage natural language descriptors as a means of further simplifying the animation authoring process, allowing other members of the production team (perhaps with no animation experience) to structure a given speech file in

terms of its overarching narrative prior to dedicated animation work commencing.

4.2 Overlay Assets

In developing the prototype, a range of image assets from various robot characters were created conforming to the skeletal shape defined for the avatar (section 4.1). Code was then added to the avatar player to allow an asset to be swapped at runtime- either as part of the configuration of the avatar player (in the case of instructional content) or in response to user selections during the interaction phase of the game (Figure 4):

Figure 4: Runtime asset selection during user configuration. *The user is provided with a list of possible body parts taken from the 8 available robot characters. When a selection is made the avatar skeleton is updated by loading in that asset, while also triggering an interaction phrase (section 3) relating to the body part.*

This aspect of avatar functionality was considered as having great potential by production partners, notably in relation to highly popular childrens gaming sites such as '*Club Penguin*' [13], that allow users to dress their characters within the game. In the case of the robot avatar, the ability to also interact with the user during this stage provides an added degree of functionality over existing methods.

5   Animation Authoring

The authoring process aims to separate the information conveyed by a particular speech asset from the final avatar and character used to deliver it. In the current prototype, the Robbie avatar can perform various movements and gestures based on the persona conceived for this character. The authoring process allows the producer to assign gestures and movements to events within a real speech file, and the rendering of those events is then performed independently by the avatar at runtime.

5.1 Speech Analysis

Emotional speech analysis work in the SALERO project has led to the construction of high quality, natural emotional speech corpora [14, 15]. Analysis of this corpus is performed using a patent pending speech analysis framework that defines various parameters in speech including: pitch data [16] and contour; intensity data [17] and contour; voice quality [18] and emotional dimensions [19]. The analysis framework outputs data in W3C standard SMIL format and has already been integrated fully with the IMDI metadata standard [20] in the aforementioned speech corpora. The speech analysis framework is used in the authoring tool developed for the animation of speech assets using the avatars described in this paper.

Each speech asset used (in any given language) during the production process must be authored to assign gestures and movements to the events in that asset, so that it may subsequently be used to drive a given avatar. The authoring tool for this prototype was developed using Adobe Air, a cross platform desktop application launcher (Figure 5):
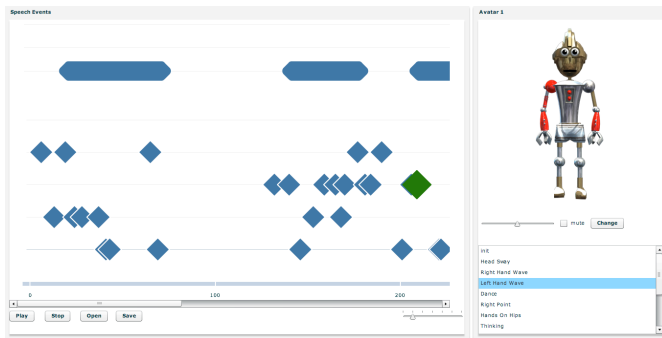

Figure 5: The Vowel Builder character animation authoring tool. *The tool allows the user to quickly allocate gestures and movements to speech events. Once gestures have been allocated, the output animation sequence is stored in the content database for rendering by the avatar player.*

The tool builds a sequence of movements and gestures and outputs them in ECMAScript (E4X) format XML [21] for storage in the content database. The use of E4X allows timing and gesture information to be exchanged easily between different rendering outputs (section 6) via the tool. A 'dope-sheet' version of the movement sequence can also be output from the tool for use in animation with 3DSMax, Maya and Blender, allowing much higher quality (though non-interactive) content to be produced. Ongoing work into the development of suitable scripts or plugins for use with 3DSMax and Maya aims to provide the animator with a pre-keyed input to accompany the relevant speech file, thus provide further savings in the production workflow.

5.2 Content storage

The delivery of interactive, multilingual content online requires a suitable strategy for the storage and management of the media assets involved. As indicated by the production

workflow (section 2.4) the authored speech assets, avatars and overlay assets must all be stored for delivery to the avatar player. By using the speech analysis framework employed in other speech corpora research carried out by the team, the entire avatar production process can leverage the capabilities of other online corpora developments such as metadata querying and asset visualisation [22]. In addition, results from emotional analysis research [23] performed using assets from these other corpora can also be incorporated into the authoring and rendering stages of the avatar production, allowing aspects of mood and emotion to be considered as variables relating to gesture and movement. As this paper details a prototype system for avatar delivery in online media production, work has been focussed on the development of the workflows and methods required prior to beta testing. For this reason, the organisation and manipulation of large number of assets and avatars has not been implemented in the current system. Future work will consider how best to implement a scalable online content management system for speech and animation content relating to avatar production.

6   Avatar Player

The current version of the avatar player is written in Adobe Flex, and is designed solely for online use. The player takes input of an MP3 speech audio file and an XML animation sequence file and synchronises the playback of avatar movements and gestures to the MP3 file at output. Additional eye blinks, eyebrow movements (not used with the Robbie avatar) and lip movements are also driven by the animation sequence file, though no authoring of these movements is required by the current prototype.

Although an avatar player has been built in Flex, other rendering engines have also been investigated. High quality renders from 3DSMax and Maya can be delivered using alpha channel punch-outs and streamed over the web via Flash Video (FLV). In this case, the use of dope sheets (section 5) relating to the movement sequences is currently used to determine the timing of keyframes. Having said this, content developed in this manner has no provision for user interaction at time of writing. Future versions of the avatar player may consider how an FLV media controller could be implemented to render high quality avatar movements, though initial tests of this method proved difficult due to the need for caching of large numbers of FLV files at runtime. While such a player is potentially possible from a technical perspective, the practical limitations of user experience criteria dictate that download times of over 30 seconds for an in-game avatar would not be acceptable.

7   Conclusions and Future Work

The avatar system detailed in this paper is currently in beta testing as part of the Peppers Ghost Productions 'My Tiny Planets' website. Testing aims to determine the capabilities of the system for reuse of avatars and assets, multilingual production and interactive gameplay. In addition, testing is

also being performed relating to the robustness and scalabilty of the system in a live context, to better assess its potential compared to existing content rendering methods (e.g. Shockwave and 3DSMax) employed within the site. The outcomes of testing will indicate where the system can be improved in terms of its user interaction and production flexibility (e.g. can more characters be added easily), while also defining the practicality of the avatar player as a scalable online delivery mechanism.

Development of the methods and tools used by this system is very much a work in progress, aiming to indicate the potential of the system for use in current and future productions. The development of rule based narratives for avatar interaction will provide a much richer interface for engagement with the user. Although limited in its current form, a more flexible combination of phrases and statements would allow the avatar to engage far more intelligently with the user. Although never as flexible as an open domain speech synthesis system, a wider range of interaction options is perhaps a more useful solution in online gaming- notably where the purpose of interaction is to inform rather than simply engage.

The creation of more avatars with a wider range of skeletal shapes and physical capabilities will also allow the flexibility of the system to be increased. The implementation of a simple robot biped for prototyping indicates the initial possibilities of an avatar, and ideas suggested during production for other skeletal types will be considered in full during future stages of production. In conjunction with this, the definition of natural language descriptors for gestures and movements (perhaps as keywords) will provide a more comprehensive delineation between authoring and rendering. A truly intelligent object should be capable of operation in a wider context, eventually aiming to possess the capabilities of human avatars in current productions. The output format of the avatar will also be considered relative to higher resolution tools such as 3DSMax and Maya. Production overhead and lack of interactivity influenced the implementation of the player using Adobe Flex, but the emergence of online gaming tools such as the Unity3D engine suggest far greater possibilities for avatar development. By developing avatars in such tools, the complexity and subtlety of the avatars behaviour can be increased considerably. This will ideally lead to more fully embodied conversational agents and avatars that can be employed in an online context, leveraging the capabilities of the current system for reuse, multilingual output and interactivity with greater granularity of movement and gesture.

Acknowledgements

References

[1.] Cassell, J.: Embodied Conversational Agents. MIT Press (2000)

[2.] Dehn, D.M., et al.: The impact of animated interface agents: a review of empirical research. International Journal of Human-Computer Studies **52** (2000) 1-22

[3.] Meadows, M.S.: I, Avatar: The Culture and Consequences of Having a Second Life New Riders Press (2008)

[4.] André, E., et al.: The Automated Design of Believable Dialogues for Animated Presentation Teams. In: Cassell, J., et al. (eds.): Embodied Conversational Agents. MIT Press (2000) 220-255

[5.] André, E., et al.: Autobriefer: A system for authoring narrated briefings. In: Stock, O., et al. (eds.): Multimodal Intelligent Information Presentation. Springer (2005) 143-158

[6.] Cassell, J., et al.: BEAT: the Behavior Expression Animation Toolkit. Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM (2001) 477-486

[7.] Piwek, P., et al.: T2D: Generating dialogues between virtual agents automatically from text.: 7th International Conference on Intelligent Virtual Agents (IVA'07). Springer LNCS, Paris, France (2007) 161-174

[8.] Morel, B.: Living Actor Avatars arrive on the scene, 2009, www.cantoche.com

[9.] Atzwanger, K., et al.: New Aspects of Human Ethology (Recent Advances in Phytochemistry). Springer (1997)

[10.] Goodman, C.: My TIny Planets, 2009, http://www.mytinyplanets.com

[11.] Haas, W., et al.: SALERO - Semantic Audiovisual Entertainment Reusable Objects. International Conference on Semantic and Digital Media Technologies (SAMT), Athens, Greece (2006)

[12.] McAuley, J., et al.: New Media: a narrative approach to content annotation. Irish Media Research Network Conference, IMRN 2008, Maynooth, Ireland (2008)

[13.] Disney: Club Penguin, 2009, http://www.clubpenguin.com/

[14.] Cullen, C., et al.: Emotional speech corpus construction, annotation and distribution. The sixth international conference on Language Resources and Evaluation, LREC 2008, Marrakech, Morocco (2008)

[15.] Cullen, C., et al.: Emotional Speech Corpora for Analysis and Media Production. 3rd International Conference on Semantic and Digital Media Technologies, SAMT 2008, Koblenz, Germany (2008)

[16.] Murray, I.R., et al.: Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. Journal of the Acoustic Society of America **93** (1993) 1097-1108

[17.] Schröder, M., et al.: The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. International Journal of Speech Technology (2003) 365-377

[18.] Hammarberg, B., et al.: Perceptual and acoustic correlates of abnormal voice qualities. Acta Otolargngologica **90** (1980) 441-451

[19.] Cowie, R., et al.: Describing the emotional states that are expressed in speech. Speech Communication Special Issue on Speech and Emotion **40** (2003) 5-32

[20.] ISLE: IMDI (ISLE Metadata Initiative), Metadata Elements for Session Descriptions, 2003, http://www.mpi.nl/IMDI/Schema/IMDI

[21.] ECMA: Standard ECMA-357: ECMAScript for XML (E4X) Specification, 2nd edition, 2005, http://www.ecma-international.org/publications/standards/Ecma-357.htm

[22.] Cullen, C., et al.: Metadata Visualisation Techniques for Emotional Speech Corpora. Second International Workshop on Adaptive Information Retrieval (AIR 2008), London, Englahd (2008)

[23.] Kousidis, S., et al.: Towards measuring continuous acoustic feature convergence in unconstrained spoken dialogues. Interspeech 2008, Brisbane, Australia (2008)