Character User Interfaces for Commercial Applications

Patricia Griffin
Headpedal, Inc.
121 Second Street, Fifth Floor
San Francisco, CA 94105
415-778-6173
pgriffin@headpedal.com

Peter Hodgson
Headpedal, Inc.
121 Second Street, Fifth Floor
San Francisco, CA 94105
415-778-6166
phodgson@headpedal.com

Scott Prevost
Headpedal, Inc.
121 Second Street, Fifth Floor
San Francisco, CA 94105
415-778-6167
sprevost@headpedal.com

ABSTRACT

Headpedal is a software company that specializes in the production of character interfaces for the web. The goal of this paper is to put forward our vision of Character User Interfaces (CHUI™s), present real-world examples culled from experience with our customers, and to describe a production process that uses state-of-the-art technology wherever possible, but also takes into account real-world constraints. Our hope is that this paper will inform researchers about the state-of-the-market for character interfaces, and assist them in judiciously applying their efforts to the development of technologies for individualizing characters.

1. INTRODUCTION

In this paper, we present Headpedal's vision of a Character User Interface and outline a methodology and procedures for creating these unique and powerful interfaces. The perceived role and personality of a character are affected by a number of elements: what the character says, the information presented by the character, the timing of responses, the character's task, and the appearance and speech of the character. All of these elements contribute to a perception of the character's individuality by users.

1.1 Character User Interfaces (CHUI™)

We define Character User Interfaces to include not only individualized characters, but also the context in which the character sits, and the interaction space in which the character can operate (following [3]). The context includes traditional Graphical User Interface (GUI) elements, such as graphics, text, buttons, and forms. The interaction space includes the access to and use of certain variables, such as user profiles and session state, and the logic by which scene transitions and character performances are chosen based on those variables and user inputs. All of these elements are part of what make up the CHUI and offer opportunities to individualize the CHUI. Of course, the individualization of the character itself within a CHUI is often paramount to achieving our clients' goals, particularly in the area of promoting their marketing brand. But it is the CHUI in its entirety that we take into account when considering notions of individualization.

1.2 Domain of Headpedal CHUIs

Headpedal presently focuses on CHUIs deployed on the web for commercial uses, primarily in the areas of online marketing, sales and customer service. Our principal aim is to deploy characters as an element of complete human-computer interfaces which satisfy consumer needs (for information, assistance, etc.) while also adhering to strict business strategies imposed by our clients. In general, our clients want to deploy CHUIs for two reasons: to differentiate their brand and their customer interface from their competitors', and to assist their customers with specific customer interaction tasks, such as paying their bill online, learning about a product, or registering for a service. To achieve the desired results we must develop unique, individualized CHUIs specifically tied to these business strategies. Unlike many research environments, we must also define efficient and cost-effective production methods.

1.3 Our view of CHUIs and Individualization

For our purposes, the requirements for individualizing a CHUI are based on context, and are not the same for all clients. In a purely marketing/branding context, it is generally necessary to make the character itself as individual as possible by focusing on appearance, gestures, expressions, dialogue style, voice and personality markers. Even so, it is impossible to develop the character outside of a broader context that includes important notions (at least to our customers) such as brand attributes (e.g. "trustworthy", "on-time", etc.) and the role the character plays in presenting the marketing message.

In a sales or customer service context, individualization of the character within the CHUI is often less important than other factors that differentiate CHUIs. For example, two characters who look much the same and have similar voices may carry out very different roles and sit on entirely different stages. In some cases, users may view these characters as individuals, despite their obvious similarities in appearance. In these cases, it is the interaction model and dialogue of the CHUI that makes one character a salesperson and another a customer support agent.

Table 1: Individualizing Elements for CHUIs

Appearance (gender, facial features, clothing, props, etc.)
Facial expressions (smiles, eyebrow raises, blinks, etc.)
Gesture sets (iconic, metaphoric, beat gestures, etc.)
Poses (frontal, ¾view, etc.)
Framing (full body, head only, head and torso, etc.)
GUI Attributes (colors, fonts, illustration styles, etc.)
GUI Elements (buttons, menus, forms, text, graphics, etc.)
Interaction Elements (logic, character memory, etc.)
Interaction Structure (decision trees, mixed initiative, etc.)

In short, all of the elements in Table 1 may serve to individualize characters for our purposes. It is the specific business context and strategies that favor some over the others for particular deployments.

Another important factor in determining how characters can be individualized is the state-of-art in character production technologies, the relative cost of various technologies, and the

market's appetite for certain types of technologies. For example, algorithms that automatically produce gestures for animated characters [2] are available and improving in performance, and are very helpful in creating characters with individualized gesture spaces, but also require extensive, domain-specific knowledge engineering. Rather than employing this type of technology today, Headpedal has chosen to develop tools that allow unskilled animators to quickly create and perfect animations by dropping gesture markers into an animation timeline using the Headpedal CHUI Editor. On the other hand, lip-sync and certain kinds of facial expressions and head movements [7] are generated automatically from a speech audio file and transcription text.

The remainder of this paper presents the objects and methods that Headpedal employs for creating unique CHUIs that solve specific business problems for our clients. In Section 2, we present the elements of a CHUI. Section 3 details our development process. For the workshop presentation, we will also present a case study centering on the development of a CHUI for a major pharmaceutical client.

2. CHARACTER USER INTERFACES: The Integration and Synchronization of Interface Elements and Character Performances

Unlike other commercial character interface approaches, Headpedal focuses on structured, task-oriented interactions driven by constrained user inputs rather than natural language query or chatterbot applications. Headpedal Character User Interfaces require the integration and synchronization of interface elements, dynamic content and character performances. The focal point of this orchestration is the interaction model. The interaction model encapsulates business goals, aims, and data structures, and executes the functions necessary to address end-user needs within a given task domain. The interaction model is a web of CHUI scenes, or states, connected by logic determined by business rules, social rules and the task domain knowledge. Based on the user's actions and profile, the interaction model executes the appropriate functionality to advance to another scene or state

Figure 1: Wait-Animation Scene

A scene or state in the interaction model is composed of both the character's performance and the content displayed within the graphical interface. Each scene has a timed duration equal to the utterance or performance of the character. For example, in the "Wait" scene in Figure 1, the character is in a wait-animation loop, and the user can browse the interface content and determine his or her interaction options. The character's wait-animation and all graphical elements comprise one scene (state) in the interaction model.

In the "Info/Account" scene in Figure 2, employed by a wireless telecommunications service, the character instructs the user to select any of the informational markers and alerts the user that they have an unpaid balance on their account.

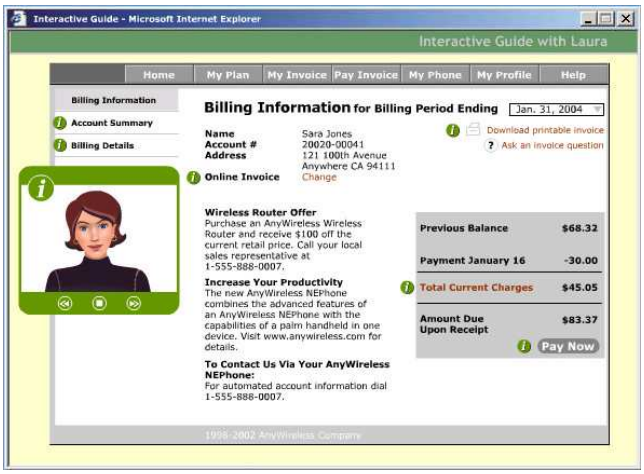

Figure 2: Info/Account Scene

3. CHUI DEVELOPMENT PROCESS

This section describes Headpedal's process for developing Character User Interfaces for our clients. The process starts with a CHUI Design phase, described in Section 3.1, which results in a structural framework and a skeletal interaction model. Section Section 3.2 describes the CHUI Development phase, which includes Character Design, GUI Design and an exhaustive Interaction Design. Finally, in the Production phase (Section 3.3), all of the assets are generated and assembled into the completed CHUI.

3.1 CHUI Design

For commercialized Character User Interfaces, the factors employed to create the individualism are largely determined based on high-level business goals and strategies. Headpedal deploys CHUIs in customer service, sales and marketing domains. Each of these domains requires a slightly different focus on the various individuating factors. In general, CHUIs designed for marketing purposes require a strongly individualized character—one that represents the brand, and won't be confused with the competition. For example, in Figure 3, the character was designed to represent strict brand qualities (trustworthy, scientific, enthusiastic, comfortable) while also fitting in with corporate guidelines for visual design style, in this case focusing on nature, science and people.

Figure 3: A Character Representing Brand

In sales and marketing contexts, where consumers are accustomed to dealing with "generic", often face-less salespeople and customer service representatives, the individualism of the character may be less important, but the character's performance and other aspects of the CHUI (e.g. interaction model, GUI elements) may require more individualization to properly convey the intended role of the character. For example, in Figure 4 and Figure 5, the two characters have the same general appearance, the same poses, the same gestures and the same voice, but are differentiated by their role, in one case a customer service assistant, and in the other a salesperson.

Figure 4: Character in a Customer Service Role

Figure 5: Similar Character in a Sales Role

Given the CHUI domain (customer service, sales or marketing), we start the CHUI design process by clearly specifying the goals of the client for the CHUI. This specification often takes the form of a Project Brief supplied by the client, or developed by Headpedal. The level of detail in the Project Brief varies greatly. In some cases, a client will supply detailed requirements. For example, the brief may specify a list of product features to be explained to consumers. In many cases, however, Headpedal must analyze client data to develop a specification that addresses high-level client goals. For example, for a client wishing to decrease costly customer service call volume, we would analyze call-center history records to determine call patterns, and the most commonly reported questions or issues. That analysis, combined with data culled from human customer service representatives, allows us to determine specific goals for the CHUI. For one major telecommunications client we found that relatively few issues comprised the vast majority of the calls received by the call center. By educating the consumer about these issues, we can eliminate a large number of calls, allowing human agents to deal with the special cases.

Because we are creating social interfaces, we must pay attention not only to our clients' business goals and strategies, but also to the goals and needs of the end-user. Often this involves trade-offs in the Project Brief. For example, our clients may wish to push certain extraneous information to their customers for marketing purposes, but the customers may prefer a more streamlined interaction.

Once the Project Brief has been specified, we identify the following items:

- Character's role in the CHUI (e.g. assistant, salesperson, tutor, store clerk, etc.)
- Use-case scenarios (examples of how the user might interact with the character and the graphical user interface)
- Main character functions and the value-add of those functions (e.g. submitting an order)
- Key variables (e.g. user profile information, account information, inter- and intra-session user information)

From these items, we develop storyboards of task domain and character interactions for discussion with our client. After several storyboard iterations, we define a skeletal Interaction Model, which contains the main scenes in the interaction, and the connections between the scenes, as illustrated in Figure 6. In developing the skeletal interaction model, we also determine the main logic of the interaction—the branching of the scenes in the interaction model based on user inputs and key variables.

3.2 CHUI Development

Once the structural framework for the CHUI has been designed, three separate design tracks can proceed in parallel: GUI Design, Character Design, and Interaction Design. In this paper we concentrate on the latter two design tracks since, with use case diagrams and UI wire-frames in hand, the GUI design work is fairly straightforward and can proceed independent of the Character and Interaction Design.

3.2.1 Character Design and Development

For our clients requiring a highly individualized (i.e. branded) character, Headpedal works with their marketing and branding teams to design the character. There are several factors at the outset of the design process that are fixed and which set certain constraints on the individualization of the character. These include the client's brand attributes (e.g. "trustworthy"), brand design guidelines (e.g. color palette) and intended market demographics.

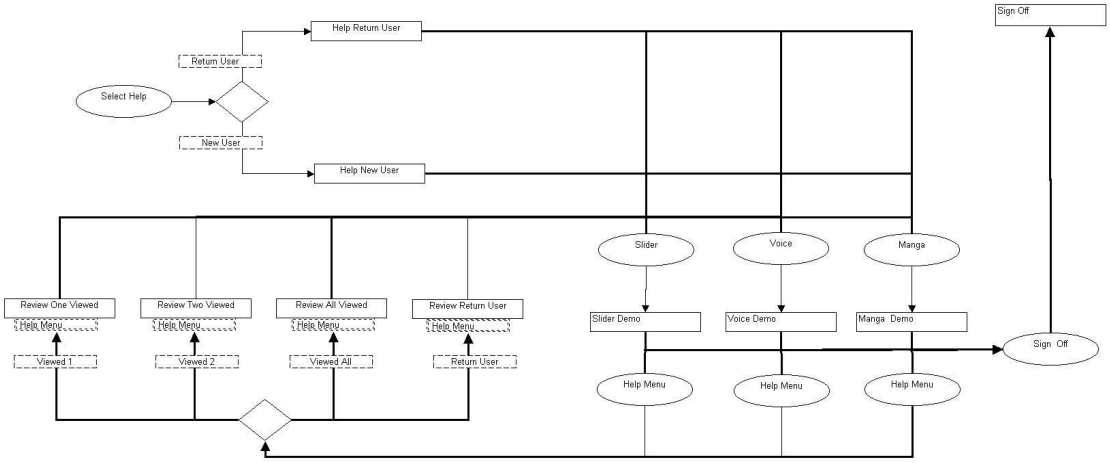

Figure 6: Skeletal Interaction Model

Given these constraints, Headpedal produces a Character Brief that describes the character's role, physical description, style and "back-story" (a description of the character's supposed history). This brief is sent to a number of illustrators who, with some artistic license, develop sets of illustrations for discussion with the client. After several iterations, a final character illustration is chosen. Once the illustration is chosen, the principle elements involved in individuating the character include gestures, facial expressions, voice, and speech characteristics.

From the Character Brief and illustration, we proceed with the design of the character's visual elements by developing a set of character model sheets. These model sheets, drawn by illustrators and animators, help to define the complete set of character poses (front view, ¾ view, side view), character framing possibilities (half-body, head and torso, etc.), character gestures and facial expressions. Once the model sheets are complete and approved by the client, we proceed with the development of the Headpedal Character Engine.

The Headpedal Character Engine is a Macromedia Flash-based object that can be embedded in a web page or another Flash application. We employ Flash rather than other rendering technologies because of the large installed base among PC users, which avoids the necessity of downloading an additional browser plug-in for most users. The character engine is a compiled Flash SWF file that contains all the animation components to render any character gesture, facial expression or pose. MP3 audio files, annotated with gesture and expression identifiers and timings, can be streamed to the engine to create on-the-fly character animations. We call theses annotated audio files 'Dialogue Score Files'. Once a character engine has been developed, Headpedal creates an engine viewer, as shown in Figure 7, which allows one to easily view the range of gestures and expressions possible for the character.

The voice of the character is in many ways just as important as the visual components. While we have used synthetic speech for some applications, most of our clients are unsatisfied with the current state-of-the-art in speech synthesis, and prefer to use professional voice talent for dialogue recordings. This inhibits our ability to create dynamically changing dialogue, but creates a sense of believability in the character that is difficult to achieve with TTS. It has also been shown that TTS is less comprehensible. [5] We must therefore create scenes in the interaction model that allow dynamic information (e.g. account balance) to appear in the graphical elements, while the dialogue elements remain static. This is not to say, however, that the dialogue cannot be rich and varied, and that the interaction model cannot take into account the discourse history in selecting which piece of dialogue to present in a given scene.

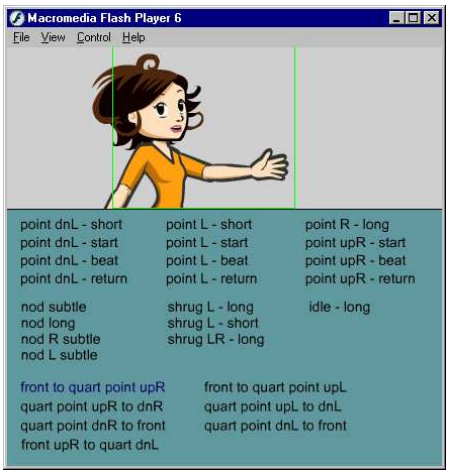

Figure 7: Headpedal Character Engine Viewer

The voice for the character is selected in a voice casting process similar to casting for radio advertisements and voice-overs. We begin by specifying a set of speech characteristics for the character consistent with the Character Brief and the character's intended role, including elements such as speaking style (e.g. formal v. informal), accent (e.g. generic American newscaster), vocabulary (e.g. scientific), and tone towards the intended audience (e.g. pedagogical). The Character Brief, speech characteristics and illustrations are sent to a talent agency, which selects appropriate voices for the casting process. Casting is an iterative process that often involves several reviews with the client before a final voice is selected.

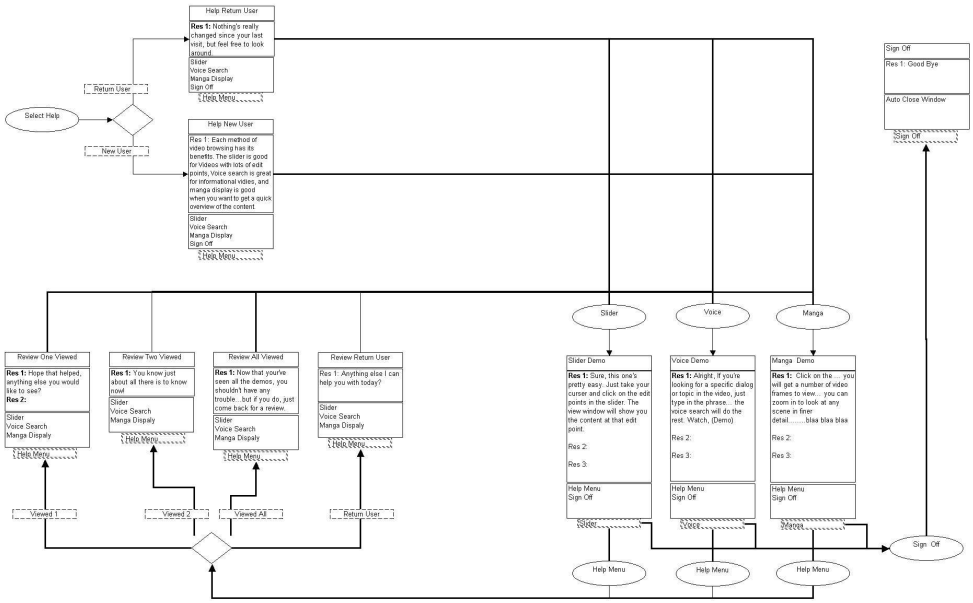

Figure 8: Interaction Model with Intent Dialogue

Note that in traditional character animation productions, the voice is often selected before the character's appearance is developed. We find that because the web is still considered a primarily visual medium, our clients are more inclined to want the visual aspects of the character design to precede the voice selection.

3.2.2 Interaction Design and Development

Each and every possible state of the system is modeled in the Interaction Model. This network of states is painstakingly mapped out, and "intent dialogue" for each character scene is written. Intent dialogue is dialogue written by the interaction designer, and only conveys the rough intention of the character rather than the precise linguistic form for realizing it (e.g. "Greet Returning Figure 8. Later, a dialogue writer utilizes the intent dialogue, along with the character profile and speech characteristics specification, to script the appropriate, complete dialogue for the character.

As stated earlier, the Interaction Model is a web of interaction scenes (states/nodes) connected by application logic. The application logic utilizes the end-user actions and profile variables to choose the correct character performance. Individualization is accomplished in a highly personalized manner by authoring states directly targeting aspects of the end-user's profile. For example, the end-user may subscribe to 'Plan-A' with a cell phone service. If the end-user's usage pattern is such that 'Plan-B' would be more economical, the character would inform the user of Plan-B and direct him or her to the appropriate location on the website to change plans.

Once the interaction model is fully developed, we proceed with usability testing, often refining the model in response to user and client input. Finally, we develop technical specifications for the back-end integration (e.g. accessing user profile variables) necessary to implement the interaction model.

3.3 CHUI Production

The final production of a CHUI involves the technical development for realizing the interaction model logic, including building back-end connections, and encoding scene transition logic in XML files that serve as input to the Headpedal Business Rules Engine. We omit a detailed discussion of these aspects of production, and instead focus here on the production of the character and other visual elements.

3.3.1 Graphical Interface Production

Headpedal has created a CHUI architecture designed to accept a wide variety of graphical user interface content. The CHUI platform contains a generalized content layout engine that allows independent production and development of graphical assets. Graphical assets are displayed and arranged according to the layout engine's specifications to create each unique graphical state' of the CHUI application. The synchronization of interface content with character performances thus creates a unified and consistent interactive experience.

3.3.2 Automatic Character Animation Production

As described in Section 3.2.1, Headpedal develops a unique Character Engine for each CHUI application. During a typical interaction the end-user will perform some action to which the CHUI reacts, by presenting new content and enacting a performance from the character. The character performance is executed by sending the appropriate annotated Dialogue Score, as determined by the interaction model, to the Character Engine. The Character Engine plays the Dialogue Score, puppeteering the correct character gestures and lip shapes in synchrony with the recorded audio speech. Thus for each state or scene, an annotated score must be created.

Creation of the annotated score files is done partially automatically and partially by hand. Lip-sync is generated automatically by our Viseme Annotation Tool (VAT), which takes

an audio file and a text transcription of the contents of the audio file, and produces a sequence of visemes and timing data. The VAT can be configured with a variety of possible phoneme-to-viseme mappings. Some types of animation tend to look better when lip-sync constraints are loosened. For example, abstract cartoon-like characters may require only a few visemes, and may look more believable when only key phonemes cause a transition from one viseme to the next [6]. The VAT can be configured to adapt to such characters.

The VAT works by first defining a constrained speech recognition grammar that recognizes only the transcription of the given audio file. That grammar and audio file are then fed to a speech recognizer that produces phoneme and timing data. This technique provides acceptable results for utterances under 15 words in length. We are currently exploring other phoneme detection tools, such as [4] for characters that require more accurate lip-sync.

Certain types of gestures are also produced automatically by Headpedal's Gesture Annotation Tool (GAT). This tool applies simple rules, reminiscent of the rules employed by [1] and [7] for gestures associated with utterance- and discourse-level phenomena. A Character Profile Editor allows certain types of events, such as pitch peaks, key phrases, beginning-of-turn and end-of-turn to be probabilistically mapped to gesture and expression events for a particular character. For example, using the profile editor, we might specify that a particular character raises her eyebrows on pitch peak events with 30% frequency, slightly nods her head on such events with 30% frequency, and does nothing for such events the rest of the time. The GAT then analyzes the audio file to crudely determine pitch events and word boundaries, and aligns gestures according to the mapping defined in the character profile. Of course the accuracy of the algorithm for determining pitch peaks varies significantly depending on the voice and the quality of the recording. In general, however, this technique, if used judiciously, provides appropriate and seemingly well-timed gestures such as head movements and subtle facial cues.

By adjusting the viseme and profile mappings, it quite easy to individualize the behaviors of a given character. Of course, not all gestures can be generated automatically without extensive domain knowledge. The automatic approach is taken in [2], in which different types of gestures that convey semantic content (iconic, metaphoric) are produced automatically from a rich knowledge representation of the dialogue domain. Because this requires extensive knowledge engineering, we opt instead to produce such gestures by hand. Headpedal's CHUI Editor tool takes the audio file and transcription as inputs, runs them through the VAT and GAT tools and produces an annotated MP3 file which we refer to as the Score. The CHUI Editor displays the audio timeline and shows the timing of the viseme and gesture events which were automatically produced by the GAT and VAT. The author can then drag and drop other gestures by hand into the timeline, and play the resulting animation. Also, gestures and expressions that were automatically inserted into the timeline can be manually removed if warranted. Once the author is pleased with the resulting animation, the CHUI Editor saves the annotated Score file (an XML file containing an MP3 pointer), which can then be streamed to the Character Engine. While this technique is not as flexible as 3D animation techniques, it is lighter weight, requires no unwieldy download for the end-user, and is easily integrated into web pages.

3.3.3 Asset Integration & Testing

Headpedal's design, development and production process is entirely iterative. The assembly of all assets into the completed CHUI is generally accomplished in three main builds. Each build has targeted areas for technical and usability testing. The initial build focuses on architectural customization. The second build focuses on the interaction model and usability, and the third tests the application on all platforms and browsers.

4. CONCLUSIONS

We have presented Headpedal's vision of Character User Interfaces (CHUIs), and discussed the various aspects that contribute to individualization of CHUIs across various commercial domains. In general, for business applications focused on marketing and branding, we individualize CHUIs by individualizing characters—through appearance, gestures, expressions, etc. For sales and customer service applications, we individualize CHUIs through the character's function and role by crafting appropriate interaction models. In this paper, we have also described a production process that uses state-of-the-art technology wherever possible, but also takes into account real-world constraints. Our hope is that this paper will inform researchers about commercial production of character interfaces, and help them to better understand the applications of their work.

5. REFERENCES

[1] Cassell, J., C. Pelachaud, N. Badler, M. Steedman, B. Achorn, W. Becket, B. Douville, S. Prevost, and M. Stone. 1994. Animated Conversation: Rule-based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents. In *Computer Graphics*, 413-420. New York: ACM SIGGRAPH.

[2] Cassell, J., H. Vilhjálmsson, and T. Bickmore. 2001. BEAT: the Behavior Expression Animation Toolkit. *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*..

[3] Churchill, E., L. Cook, P. Hodgson, S. Prevost, and J. Sullivan. May I Help You?: Designing Embodied Conversational Agent Allies. In *Embodied Conversational Agents*, Cassell, Sullivan, Churchill, Prevost (Eds.), 64-94. Cambridge: MIT Press.

[4] Erler, Kevin and George H. Freeman. 1996. An HMM-based speech recognizer using overlapping articulatory features. *J. Acoust. Soc. Am.*, 100(4):2500-2513.

[5]] Gong, Li and Jennifer Lai. 2001. Shall we mix synthetic speech and human speech: impact on users' performance, perception, and attitude. *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*. Seattle.

[6] Griffin, P., and H. Noot. 1993. The FERSA Project for Lip-Sync Animation, *Proceedings of IMAGE'COM 93*, 111-120.

[7] Pelachaud, C. and S. Prevost. 1994. Sight and Sound: Generating Facial Expressions and Spoken Intonation from Context. In *Proc. of the 2nd ESCA/AAAI/EEEI Workshop on Speech Synthesis*, 216-192. New Paltz, NY.