Essistants

Christos Tryfonas
Sprint ATL
1 Adrian Court
Burlingame, CA 94010, USA
tryfonas@sprintlabs.com

James Schumacher *
Siftology Inc.
5829 Viramar Ave
Oakland, CA 94618, USA
transeme@yahoo.com

ABSTRACT

One of the challenges the service providers currently face is the ability to introduce a variety of services at a minimal cost and impact to their customers. These services often become personalized as more and more content becomes available. The natural progression to the service/content explosion is a seamless user interface that remains consistent across the various services and devices. The Essistant architecture attempts to provide personalized services to the end users through a seamless multi-modal user interface and a systematization of the backend services. This paper describes the overall architecture of the Essistant project. The associated video demonstrates the functionality of Essistant for a set of services that have been implemented in a lab environment. The services include video-on-demand using an automated price broker, broadcast video over IP multicast, personalized news, and horoscope, and interaction with the physical space by acting as a proxy to a robot.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation (e.g., HCI)**]: [Animations, Artificial, augmented, and virtual realities, Audio input/output, Video (e.g., tape, disk, DVI)]

General Terms

Human Factors

Keywords

Multimodal Interfaces, Multimedia Presentation, Avatars, Service Discovery

1. INTRODUCTION

The Essistant is built upon on a distributed network of services and devices. It delivers personalized content and services to users in a seamless manner. The Essistant utilizes agent technology and network-based user information to interact with the user. Similar systems have been demonstrated in the past with focus on specific applications such as assistance for real-estate and office administration [1, 2, 4].

The Essistant is about ubiquitous and pervasive interaction with common Internet and broadband services. It employs a multi-modal interface to such services as news, traffic, email, and home automation.

2. ARCHITECTURAL DESCRIPTION

The architecture of the Essistant system is depicted in Figure 1. The client side consists of the voice/face recognition components and the multimedia player that displays all the relevant media types to the end user. The server side can be split into the following distinct entities:

1. The component that handles the content/service profiles for each user.
2. The content databases that contain the associated content for every supported device/profile.
3. The Text-to-Speech (TTS) server to create the audio file/stream based on a specific text segment.
4. An Artificial Intelligence (AI) engine that performs all the necessary associations for relevant content that a user profile might be interested in.
5. The presentation layer component. Within this component, a multimedia server interacts with a closely coupled animation engine to achieve the avatar functionality and media synchronization.

2.1 Information Flow

The voice recognition component performs the initial extraction of the word tokens that are fed into the Natural Language Processing (NLP) engine on the server side. The face recognition signals the corresponding name/profile to the profile manager on the server side for activation of the corresponding content/service databases. The multimedia player presents the animated Essistant with all the associated media types based on the associated profile / device / bandwidth tuple. The multimedia player is driven by the corresponding server side component that performs all the necessary synchronization of the media sources to be used in the scene composition.

The profile manager at the Services Controller keeps track of the content needed per user profile for all the supported devices. Both the animation engine and the media server

*James Schumacher is currently affiliated with Siftology, Inc., 5829 Viramar Av, Oakland, CA 94618, USA. He can be contacted at transeme@yahoo.com. This research was done while the author was at Sprint ATL.

MM'03, November 2–8, 2003, Berkeley, California, USA.
ACM 1-58113-722-2/03/0011.

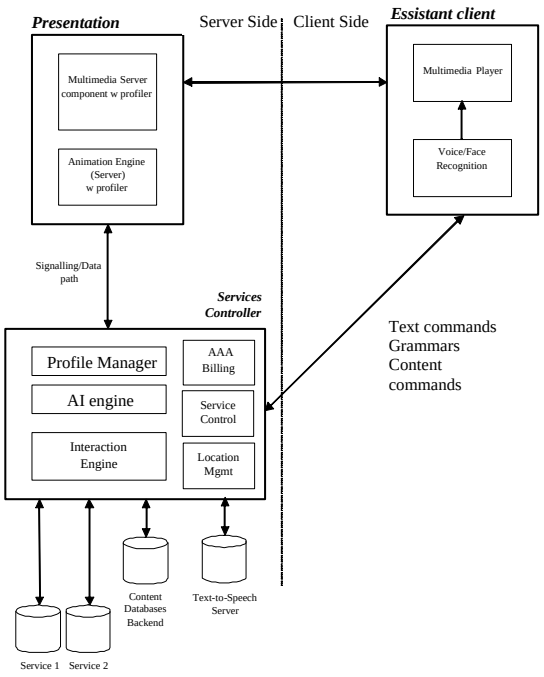

Figure 1: Generic Architecture of Essistant.

use different profilers. The animation engine profiler is responsible for the generation of the various Essistant behaviors. The corresponding user profile should be signaled by the profile manager to both the animation engine and the multimedia player through the relevant signaling path.

3. SERVICES

The architecture defined above is based on the notion of services. Every supported service in the Essistant architecture has the following structure:

- A signaling mechanism to the Services Controller component notifying it that the service is available along with other health monitors.
- A control plane with a service-specific protocol to signal the various controls to the Services Controller component.
- A context-free grammar that is used by voice recognition system and the association between the grammar actions and the commands provided by the control plane.
- A data plane that should be able to be controlled by the Services controller component using commands defined in the control plane.
- A set of scene descriptions for the presentation of the service to the client device (presentation layer).
- An association of the scene objects with the service controls as they are defined in the control plane.
- A set of interfaces to other services also featured in the system. For example, if the user watches a movie and there is an incoming call/message, this interface should define how to signal the VoD service for the incoming message.

4. DESCRIPTION OF VIDEO

The associated video, "Essistant," chronicles the capabilities of a speech enabled interface to a ubiquitous computing environment intended for a home or office. Making use of a dynamically animated personified agent, the Essistant gives a single personality to the myriad of services at the users command. The Essistant provides the home user with the ability to control the television, adjust the lights in the room, read email, check the news headlines, and even control the household robot.

The user controls the services through the personality of the animated agent. Behind the scenes a set of loosely coupled network services provide the functionality. A profiling service tracks and builds user preferences for news and traffic reports. A set-top box is enabled with a network proxy capable of receiving basic commands for television control. An 802.11b enabled robot, can likewise carry out commands received over a simple TCP protocol. And finally as the user requests to watch a movie via Video-on-Demand (VoD), the Essistant arranges for agents to search for the title, broker the best price [3], arrange for bandwidth, and then play the movie.

5. FUTURE WORK

We are hoping to see the Essistant architecture evolve into a more standardized plug-and-play service model, in which third-party providers will be able to host their services and applications, and have them presented using this new medium. Also, we need to integrate all the multimedia sources (TV, VoD, animation) into the same scene using the emerging MPEG-4 set of standards.

6. CONCLUSIONS

We showed an overview of the Essistant concept and demonstrated some of its functionality using the associated video. The description of the architecture of Essistant is by no means exhaustive and only attempts to present the information flow in the overall system architecture.

7. REFERENCES

[1] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjlmsson, and H. Yan. An architecture for embodied conversational characters. In *First Workshop on Embodied Conversational Characters*, Tahoe City, CA, 1998.

[2] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjlmsson, and H. Yan. Embodiment in conversational interfaces: Rea. In *ACM CHI 99*, Pittsburgh, PA, 1999.

[3] J. C. Straley, D. C. Keenan, M. O'Brien, and C. Tryfonas. The importance of being efficient. In *Proceedings of Int'l Conf. on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet (SSGRR) 2001*, August 2001.

[4] H. Yan and T. Selker. Context-aware office assistant. In *International Conference on Intelligent User Interfaces*, New Orleans, Louisiana, 2000.