Multimodal Interaction with Interactive Television

Aseel Ibrahim
Nokia Home Communications
Department of Computer and Information Science / Linköping University
aseel.ibrahim@nokia.com

Abstract

Interactive television provides viewers with new services and applications, it therefore requires an effective interaction mechanism. There are many design challenges to meet in interactive television and some of these challenges can be handled by offering multimodal interaction where natural language is used in combination with a remote control. The purpose of this paper is to present my ongoing work that is focused on investigating the possibility of using multimodal interactivity in order to provide the user easy interaction.

Introduction

What is interactive television? People are used to interact with computer-based system that supports communication in two directions: from user to computer and vice versa. Such systems are known as interactive systems. Today, television content broadcasts and viewers are unable to affect television viewing. However, the development of the traditional television includes adding interactivity to television content and allowing the viewer to access and interact with associated information, while watching television. This is often referred to as "interactive television" that provides viewers with new applications and services besides viewing television content.

Many of the services that today are provided by computers will also be provided by the interactive television. Sending e-mails, purchasing via e-shopping, playing games, net surfing, as well as information searching are some of the services. These services require text input, information searching and navigating. The viewer has the opportunity to search among television and Internet content. Hundreds of television program channels will be accessible for the viewers. Today, viewers select program by zapping through channels or using printed channel/program guide. This strategy will fail with the huge amount information about hundreds of channels.

Challenges that have to be faced when designing the new medium are many, but the main challenge is to provide an easy interaction mechanism in order to achieve the user intention. Multimodal interaction where natural language is used in addition to a remote control could provide the user with such easy interaction.

Multimodal Interactive Television

Eye- and hand-busy are common arguments for using speech in multimodal systems, but why use multimodal interaction in interactive television? The speech input is conducted by natural language. Today, speech in voice-controlled television is mainly used for saying commands. Using voice commands requires some cognitive effort that could be a hindrance for the user's intentions because the user has to learn these commands and to remember them in order to be able to use them. One of the design challenges is that television is an entertainment device that viewers would not have to work hard to learn and use. Therefore voice commands are not useful. On the other hand, the effort of learning and remembering is very low with natural language. For this reason natural language is suitable as an interaction mechanism. According to Oviatt (1999) there are many advantages in using multimodal interaction; multimodal interaction gives the user freedom to choose a modality that supports her to perform tasks, supports better error handling, and increases efficiency. These advantages contribute to a more efficient interaction.

Below is an example of how natural language can be used for searching television content:

User:	"Are there any movies running right now?"
TV:	"There are three movies running right now. The patriot is running on TV4, Mission Impossible 1 on TV1000, and The Matrix on ZTV."
User:	"How much have I missed from "The matrix""
TV:	"It started about 15 minutes ago."

User: "I want to see it."
TV: "I'll zap to ZTV now."

In this example only one modality is used, speech, for both input and output. This dialog example could be shortened by adding visual or textual interface. Using graphics in addition to speech gives the following scenario:

User: "Are there any movies running right now?"
Four pictures with the running movies are viewed.
User: "I'd like to watch movie one"

The pictures mentioned in the example could be four different program scenes. The graphical presentation can be realistic if the amount of information is limited. In addition, the low resolution and the long distance between the viewer and the television decrease the amount of information that can be viewed. On the other hand, using graphical presentation can increase the user's understanding. The graphical presentation can be exchanged with textual presentation, but the problems addressed with the amount of presented information are remained. The amount of the presented information is critical and influence the choice of the presentation technology.

Another issue that must be taken in consideration when providing multimodal interaction is the user's attitude. The remote control will be equipped with a microphone that can be activated by pressing a specific button. Talking to a device can be experienced as unnatural which can feel uncomfortable for the user and having a face to talk to could be a solution.

Research Proposal

The aim of my ongoing work is to investigate how multimodal interaction, natural language dialog in addition to a remote control, can be used in order to design better interaction with an interactive television.

This will be done by examining to what extent natural language can be used in order to offer the end-user a usable interface that it is easy to interact with. For this to be possible we must define usability. The focus of usability in this case is on the user's attitude as well as system efficiency. The user's attitude is about the user's satisfaction with the interactive television, the user's perception of control of the interactive television, and finally the user's perception of freedom when using the interactive television. Efficiency is about cognitive overload, error, time to perform tasks, and ease of use.

Input and output modalities and their multimodal combinations are the focus of the investigation. The output can be provided in various ways: speech, graphics, and text output. Speech output can be unsuitable because of the noise of the television, and the background noise.

Using multimodal interaction where natural language dialogue is combined with a remote control to interact with interactive television is a challenge in order to satisfy the end-user and to build a system that is both easy to use and effective. Providing multimodality increases the user's freedom to choose a suitable mode and gives the user control. In order to be able to face the challenges the following questions are interested to investigate:

1. User response: How do users feel about speaking to a television? How do their attitude effect the ability to interact?
2. Combination of output modalities: Which kind of information should be presented by text, speech, and graphics?

References

Oviatt, Sharon (1999), "*Mutual Disambiguation of Recognition Errors in a Multimodal Architecture*"; *Proceeding of the CHI 99 conference on Human factors in computing systems: the CHI is the limit*, 1999, Pages 576 - 583