Chapter 1. Introduction

Section 1.1. Problem Description

This document describes much of the research and development that was involved in a three-year project aimed at the creation of an interactive television device. The Center for Wireless Telecommunications (CWT), a research group within the Bradley Department of Electrical and Computer Engineering at Virginia Tech, was approached by a group of investors who had a new concept for interactive television. The president of Interactive Return Services, Inc. (IRS) based in Herndon, Virginia, headed the investment group and guided much of the long-range research focus. He was joined by PISA, a group of Mexican investors. Additional funds were provided by the Center for Innovative Technology (CIT), and CWT itself.

In the fall of 1994 Dr. A. A. (Louis) Beex was invited by the CWT to work on the project as an audio signal processing expert, and I began to work under Dr. Beex's supervision in the spring of 1995. The Assistant Director of the CWT, Mr. Willard W. Farley, coordinated the efforts of the various research teams, and worked very closely with our audio group in particular.

The investors had an idea for an innovative approach to interactive television. They envisioned using the audio component of a television signal to either actively carry extra information (an inserted digital signature) or to passively identify a commercial or program. The television viewer could use a special device capable of extracting and decoding the audio data to order products and services, or to request information and

coupons. The device, later dubbed the "AudioLink," could also be used to automatically collect viewing statistics as done for Nielsen-type ratings, and would also allow interactive games. Internet web-browsing capability and the detection and characterization of audio alarms would be additional options added later. A wireless radio frequency (RF) link would transmit user interaction results to local receiver cell sites, called repeaters, which would then pass the messages to host computers. These host computers would take action as appropriate, based on the information received and the user responses. The entire system was collectively called the Interactive Video and Data System (IVDS), and is diagrammed below in Figure 1.1.

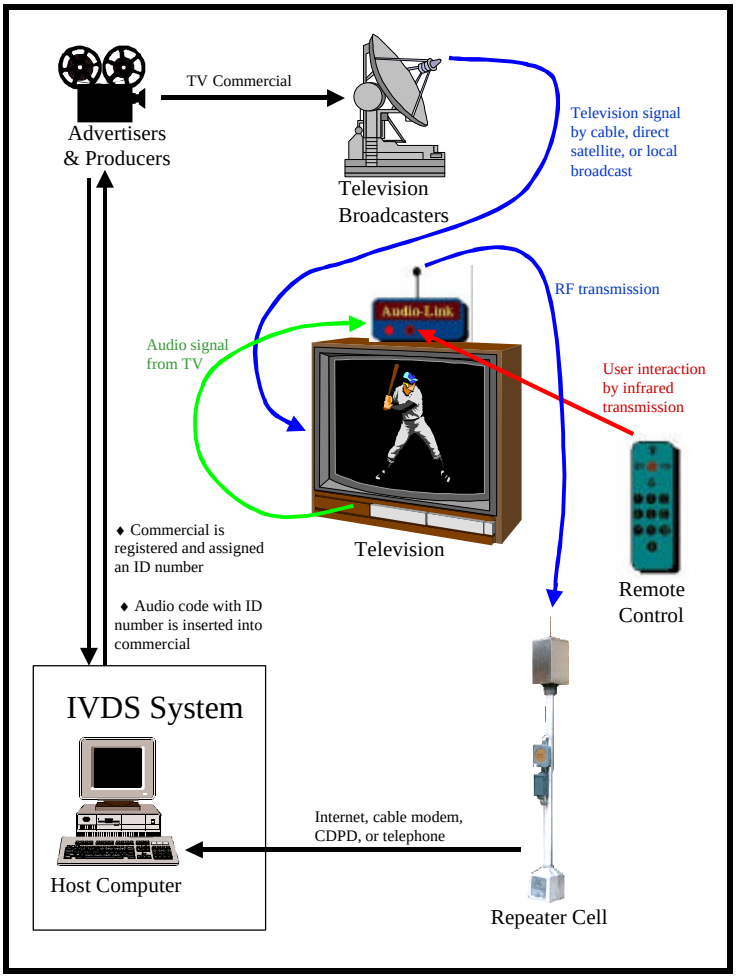

Figure 1.1 IVDS System Diagram

A series of constraints were specified for the AudioLink. The target device was to be small, inexpensive, battery-powered, and hand-held. (Originally the hand-held device with an embedded infrared transmitter doubled as a universal remote control. Late in the project the hand-held and battery-powered constraints were removed and the AudioLink became a set-top box as depicted above in Figure 1.1. This box contains an infrared receiver which decodes signals from any universal remote control.) For the mode of operation where the television audio was to be modified, the embedded digital

signatures had to be about 35 bits in length, and should not be objectionable to human observers. Four of the 35 bits would be used to specify an AudioLink command, and the remaining 31 bits would identify the particular commerical or television program. Ideally the inserted codes should be completely imperceptible. The AudioLink should be able to easily and quickly detect and extract the hidden codes, and take action based on the information contained in the codes. Furthermore, no connections to the television were allowed, and interception of the television signal before the television was not allowed. Additionally, the coding method should be relatively robust against extraneous room noise, especially human voices. The main focus of this thesis is the research and development involved with such active insertion of codes into an audio signal and the detection of such hidden codes. Other aspects of the project will be succinctly presented for completeness and clarity, but will not be covered in full detail.

As the project evolved over the course of three years, many areas of research and development were pursued concurrently. Unfortunately a sequential document such as this is not really appropriate for accurately describing such concurrent development. Thus “looping” is inevitable, and some parts of the project will need to be revisited after their initial description. As individual facets of the project are described in different contexts, appropriate cross-references will be provided wherever possible.

Section 1.2. Early Parallel Investigations

As briefly mentioned above, one possibility for achieving interactivity was to use the unmodified audio signal to passively identify a television program or commercial. If

TV audio signals could be properly characterized, they could possibly be distinguished from one another based on selected representative parameters. For this scenario it was imagined that a user would be watching a given television commercial, and would click a button on his or her remote control to perhaps order the product, request additional product information, or request a coupon. Prompted by the remote control keypress, the AudioLink would begin sampling the audio signal and characterize it in some manner. The parameterized representation would be transmitted by the AudioLink, along with the user's response and identification number, to a local cell site and ultimately to the host computer(s). The host computers would then be able to identify the commercial by comparing the parameterized representation of the sampled audio with values calculated a priori and stored in a database. Once the commercial was identified, the user's request could be processed as appropriate. The sponsors also desired the ability to determine the moment the keypress occurred within the commercial, to whatever degree of accuracy was possible. Such information would be useful in circumstances where multiple items were advertised in a given commercial, or where special privileges were bestowed upon the first users to respond (i.e., during a game or contest).

Dr. Beex and a graduate student named Gregory Sheets investigated such passive use of the audio during the first year of the project. Promising results were achieved by using linear prediction (LP) coefficients, reflection coefficients, or their cepstral coefficient counterparts [1]. Audio segments of 10 ms duration were extracted from the audio at 1-second intervals, beginning at the moment of the keypress. The audio segments were analyzed and appropriate LP (or other) characterizations were computed.

Quantized versions of the characterizations were used in an attempt to identify the commercial and timing information from database entries computed a priori.

Since the parametric representations were to be used to uniquely identify a given commercial, as well as the temporal location of the keypress during the commercial, much attention was given to the choice of appropriate characterizations. An ideal representation would provide a high degree of correlation with the database for the correct commercial and for the correct temporal offset. When compared with the database entries for other commercials, and for other temporal offsets, an ideal representation would produce low correlations. Furthermore, since the parameterized representations had to be transmitted to the host computer in a relatively small number of bits, the parameters most relevant for identification had to be chosen and quantized, and the least relevant had to be mostly if not completely ignored. Much research effort was directed at choosing which of the parameters are most relevant for discrimination, and how the discrimination is affected by parameter quantization.

After an initial investigation several possible solutions were analyzed in detail. While the extraction of audio and the calculation of parameters at the AudioLink was relatively simple and doable, the database lookup proved to be too computationally demanding. The cross-correlation calculations were too intensive, and became more so as the number of candidate commercials grew. Thus, as the system became more popular and more commercials were added to the database, more correlations would have to be calculated, and significantly more processing power would need to be added. Furthermore, as the commercial numbers grew it became more and more difficult to

identify a particular commercial from a group of "similar" ones. The threshold region between correct and incorrect matches grew smaller, making proper identification more difficult. From a business standpoint the drawbacks were deemed unacceptable, and the sponsors abandoned the passive use of the television audio. This thesis will not cover the passive use of the audio in any additional detail, but the interested reader can refer to Gregory Sheets' M.S. thesis [1] for more information.

During the first year of the project the target device was to be battery-powered and hand-held, and would double as a universal remote control. Because of the power consumed by the digital signal processor (DSP) and the RF transmission, the battery life was quite limited. Eventually, as will be described later, the sponsors heeded the suggestions of most of the research teams and changed the target to a set-top unit powered from an AC adapter. Since power was no longer an issue, the DSP could operate continuously and the audio could be sampled continuously. This allowed the option of listening for audio alarm signals. If audio alarms such as smoke, carbon monoxide, and intruder detectors could be accurately detected and identified, the host computers could be alerted to the presence of such alarms and action could be taken as necessary.

The first step in pursuing the alarm detection capability was to identify the audio characteristics of various alarms and determine which, if any, could be used to detect and uniquely identify them. Another of Dr. Beex's graduate students named Ananth Padmanabhan was tapped for this investigation. Dr. Beex and Ananth found that most common household alarms have time-frequency-amplitude characteristics that could be

used for detection and identification [2]. In fact, the audio alarms follow unofficial standards (manufacturers are working on finalizing official standards), and alarms within groups (e.g. smoke detectors) possess similar characteristics. For instance, alarms within a group have specific frequencies of tones present, and those tones would cycle on and off at a particular rate and duty cycle. By determining the frequencies present in an alarm signal, and the repeat rate and duty cycle, the alarm could be classified into its group. Although some of the implementation details of alarm detection and identification will be described later, Ananth's research will not be covered here. If more information is desired, the reader can refer to a final report [2] that presents all research and results.

As the project moved from the pure research phase toward hardware implementation, a new graduate student joined Dr. Beex's audio team. Sundar Sankaran was responsible for converting the AudioLink code written in Matlab to assembly language for the chosen digital signal processor. He also wrote the shell program to control the AudioLink, and was responsible for integrating the various software modules written by other research groups (such as the module for RF transmission). He also continued Ananth's work with the alarm detection, and investigated and implemented the infrared (IR) detection scheme for deciphering remote control keypresses.