

Chapter 9. Conclusion

The Interactive Video Data System (IVDS) project began with an initial abstract concept of achieving interactive television by transmitting hidden digital information in the audio of commercials. Over the course of three years we successfully developed such a communication method, designed and built the hardware systems to realize the application, and conducted several full-scale field tests. Researchers at universities rarely get to participate in the entire cycle of product evolution, including concept generation, research, development, testing, and marketing. We were fortunate to have been involved in every phase of the project, and especially to have our work culminate in a working prototype hardware device – a very tangible exhibit even to people outside of the engineering world.

The coding scheme we created satisfies all of the design constraints imposed by the project sponsors. By taking advantage of psychoacoustic properties, our hidden digital signature is inaudible to most human observers yet is detectable by the hardware decoder. The communication method is also robust against most extraneous room noise as well as the wow and flutter of videotape machines.

The hardware systems we designed for the application have been tested and work as intended. The triple-stage audio amplifier buffers the input signal, eliminates low frequency interference such as human voices, and boosts the filtered result to an appropriate level. The codec samples the filtered and amplified audio, and feeds it into the digital signal processor. The DSP, after applying a pre-emphasis and compensation

filter, calculates the FFTs, compensates for frequency shifts, extracts the digital signature, and verifies the result via the cyclic redundancy check. It then takes action appropriate for the command specified in the digital signature. If necessary it will verbally prompt and provide information to the user, and will decode infrared signals from a remote control. The results of interactions are transmitted by radio frequency spread spectrum to a cell site, where they are then forwarded to the host computer.

All research and development for the entire IVDS project was conducted at Virginia Tech. Dr. Beex and his graduate students in the DSP Research Lab were responsible for all of the audio signal processing and communications, the infrared signal decoding, the programming of the digital signal processor, and all of the associated hardware design and testing. This range of responsibility essentially covered the entire AudioLink device, with the exception of the spread spectrum transmitter. In fact, even the transmitter was partially our responsibility, since its components were controlled via the DSP program.

I was personally involved with almost all aspects of the design as a team member, with more emphasis on the audio signal processing and communications aspects of the project. Under Dr. Beex's supervision and guidance I created the sinusoidal keying communication method, invented the control function for synchronization and data weighting, and developed the frequency locking mechanism to compensate for tape wow and flutter. I also investigated and implemented the triplication code for error correction and the cyclic redundancy check for error detection. Furthermore I designed the audio amplifier and filter circuits which pre-process the received acoustic signals, and I

sampled, processed, and converted the voice messages to be used by the AudioLink when conveying information to the user or prompting for interaction. My direct responsibilities also included the entire code insertion process, and I inserted all the codes for the numerous demonstrations and field tests. In addition I wrote the windows program to facilitate code insertions by others.

Section 9.1. The Planned Future of the AudioLink

After the next series of field tests in both Mexico and the United States (as mentioned in Section 7.5), the first full deployment of the IVDS system is likely to be in Mexico. The PISA investors are currently searching for additional capital investment, which would permit the final stages of system development. From there they reportedly would like to conduct a full marketing test in Mexico City, and wider coverage would be added later as demand increased.

Several AudioLink enhancements are planned for the future. Bar code reading capability has already been investigated as a possibility, although work was suspended in order to concentrate more effort on the main interactive television application. A bar code reader wand would be interfaced through a port, and the DSP would be programmed to read UPC bar codes. The idea is that a user could swipe the wand across bar codes to order a product or service. The bar codes could be for products or services advertised in custom catalogs in a home-shopping application, or they could come from existing products so the user could order replacements, which would allow grocery shopping from home for instance.

Another option that was investigated but then temporarily abandoned was a keyboard interface. The AudioLink is already equipped with an infrared receiver, and the decoding of remote control signals is done in the digital signal processor. Wireless keyboards using IR signaling already exist, and could be used in conjunction with the AudioLink. This would especially be appropriate for use in browser mode, when the user is surfing the internet, and would allow the sending and receiving of electronic mail. Expansion with the keyboard interface would primarily require additional programming of the DSP, and few hardware changes are anticipated.

Section 9.2. Possible Improvements and Suggestions for Future Work

Section 9.2.1. Improved Detection Capability

If additional time had been available, methods of improving the code detection capability could have been investigated further. Such potential improvements are now left as suggestions for future work. The use of a microphone array coupled with additional signal processing could perhaps allow operation without a clip-on microphone or a direct connection by an RCA terminal. Such an enhancement might also permit the handheld device option, which was the original project goal. Adaptive beam-forming techniques could be used to focus on the television speaker as an audio signal source, and other room noises could be rejected. The introduction of additional microphones might also allow adaptive noise cancellation techniques to further reduce interference from extraneous room noises.

The analog audio amplifier and filter circuit can also be improved. In fact, a revised version has already been designed, but has not yet been fully tested or implemented. This new circuit further rejects interfering low frequencies, possesses a flatter passband with a smoother rolloff, and is less sensitive to component variations. Furthermore, it is less susceptible to circuit noise, which was a problem given the high gain in the third filter stage.

Another option that was briefly investigated but was abandoned because of impending deadlines was the use of an audio signal compressor circuit. As discussed in Section 5.2.4, such signal compression could be quite useful in eliminating clipping and in maintaining sufficient use of dynamic range during analog-to-digital conversion. If the compressor circuit parameters are chosen correctly, enhanced AudioLink detection should be possible.

Section 9.2.2. Improved Error Correction

As discussed in Section 3.1.3 the error correction capability might be improved if codes other than the triplication code are employed. One promising option is the (15,5) $t=3$ BCH code, which could be used in ten interleaved groups to maintain sufficient robustness against burst errors. Its performance would have to be evaluated empirically, since individual error statistics are difficult if not impossible to compute in this application. The move to an alternate code will require major changes to the DSP program, and so a new code should prove to be significantly better than the triplication code before it is adopted. In contrast, an enhancement that should be implemented immediately in the next revision of the AudioLink is the soft decoding of data bits, as

described in Section 4.7. Experiments have already shown that such soft decoding improves the effective signal-to-noise ratio by about five dB over hard decoding.

Section 9.2.3. Improved Signal Masking

A couple of proposed changes to the code insertion process could perhaps improve the hideability without degrading the detectability. Or, equivalently, the detectability might be improved without compromising the hideability. Regardless of from which direction one considers it, the possibility exists to place more energy in the code sinusoids and still maintain concealment. One way that this might be achieved is by adapting the sinusoidal amplitudes to conform to the power spectral density of the original audio. In other words, each sinusoid would be “responsible” for representing the original audio energy in its neighborhood, and its amplitude would be varied accordingly. Limits would have to be imposed to ensure that all sinusoidal amplitudes remained within some range of each other. By varying the amplitudes the code sinusoids might more effectively replace the original audio content, and so their presence might be better hidden.

Another possibility is to shorten the code duration below the 268.8125 ms lower limit. As explained in Section 3.2.2 this limit is imposed to ensure that at least one FFT block on the decoding end will be completely contained within the code’s duration. However, if the sinusoids have sufficient amplitude, they can successfully compete with audio energy in the same frequency band. Therefore, a shorter but slightly louder code might be better concealed due to psychoacoustic masking, and still be detectable. The

relationship between code duration, amplitude, and hideability is nonlinear, and so experiments will probably be necessary to determine what is actually possible.

Both of these options would have to be evaluated by trial and error and compared with the current process to determine the effects on hideability and detectability. In the end there will always be a tradeoff between hideability and detectability, and the IVDS managers must make a choice between detection rate and perceived audio quality. However, these two potential enhancements in the insertion process might allow a larger acceptable range of use between the two extremes.

Section 9.2.4. Automatic Selection of Code Insertion Locations

One fertile topic for future research is the automatic selection of candidate code insertion locations. Although tools like the spectrogram are useful in identifying candidate locations, the code insertion process still relies on human interpretation of the spectrogram results to deduce candidate locations, and trial and error at the candidate points. An automatic process would of course be desirable, especially since the persons inserting the codes in the future will often lack engineering training.

The time-frequency characteristics most important for masking the code sinusoids would first have to be identified. An automated program would have to compute the spectrogram (or a similar representation) and search it for promising locations to propose as candidates. It is likely that the best insertion locations would be where there is much low frequency energy present, and where there was prior strong energy present in the code region, as explained in Section 3.2. Energy at higher frequencies at the time of

insertion and energy in the code region just after the insertion time are also important, but to a somewhat lesser extent.

Perhaps the spectrogram could be treated like an image, and a template could be created to find the best candidate insertion locations. Figure 9.1 below shows an example template, similar to a non-symmetric “O,” which would indicate where favorable conditions exist for hiding the sinusoids. Note that the template is weighted, and more emphasis is placed on prior time and lower frequencies. Also, the weights decay away from the candidate insertion point, since proximity is extremely relevant both in time and frequency. This template could be slid across the spectrogram in time, and strong correlations with the time-frequency content of the audio signal would reveal promising insertion locations.

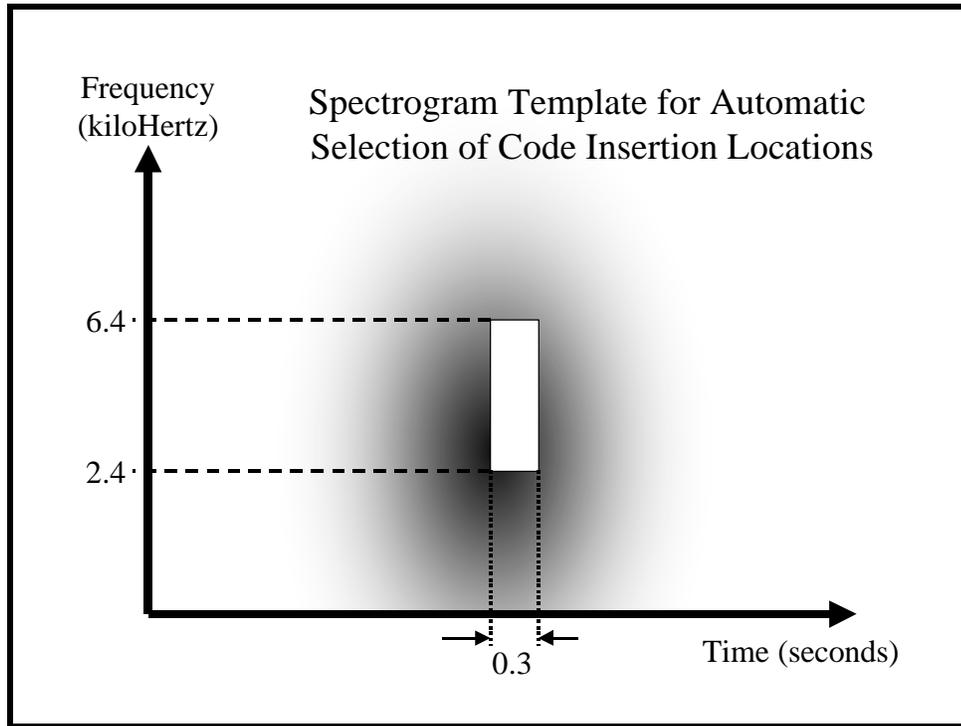


Figure 9.1 Template for Automatic Selection of Code Insertion Locations

Another factor in code hideability is the rate at which the background audio changes. This too could be introduced as a criterion for the selection of insertion points. However, final decisions should still be verified by trial and error with human observers. Experience has shown that sometimes the “best” areas from a theoretical point of view do not hide the codes well in practice. This perhaps suggests that more research is required to fine-tune the existing psychoacoustic theories.