Digital Archeology: Recovering Digital Objects from Audio Waveforms

Mark Guttenbrunner
University of Technology Vienna
October 6th, 2009

http://www.ifs.tuwien.ac.at/dp

Overview

- Introduction
- Analyzed System
- Reengineering the Waveform
- Bitstream-Formats
- Migration Tool
- Evaluation of Extraction Methods
- Evaluation Results
- Conclusions

Introduction

- First home-computers in the late 1970's / early 1980s
- Data usually saved on magnetic tapes
- No special hardware/media needed (e.g. standard audio recorder to connect the system to and audio tapes to store the data)
- Data/Software still available on tapes in private collections (maybe even archives)
- Future migration with no working specimen and/or knowledge about system impossible

How can we migrate the data without
the original system in the future?

Analyzed System

- **Philips Videopac+ G7400**
  - video game system
  - released 1983

- **C7420 Home Computer Module**
  - Microsoft Basic
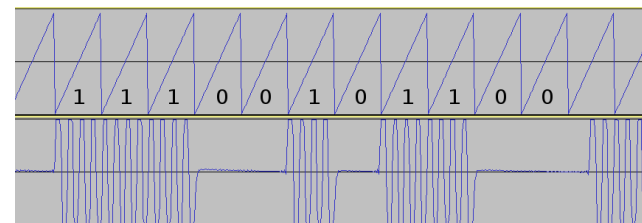 - Data save/load using compact cassettes

Reengineering the Waveform

- Data is encoded in bitstreams
- Bitstreams are encoded in analogue waveform
- Writing different test programs on the original system
- Recording digitized waveform using software Audacity

- Analyzing waveforms and changes between waveforms
- Format of stored byte:
  - sine-waves, 4.8 kHz
  - 4 waves per bit, 1200 bps
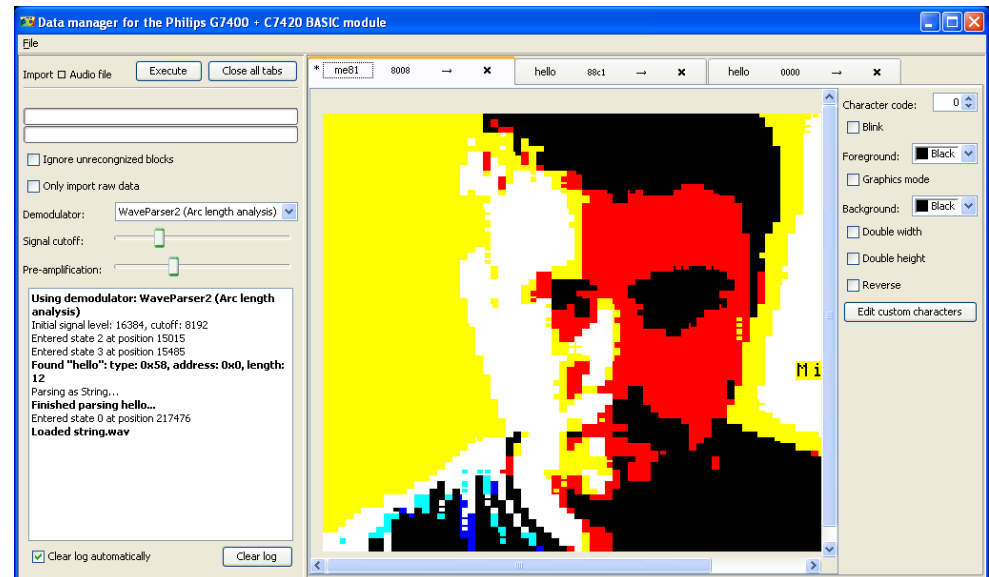 - 1 start-bit, 8 data-bits, 2 ½ stop-bits

Bitstream Formats

- ## System is able to store different file types
  - BASIC programs, screenshots, arrays, strings, raw memory dumps
- ## Writing test programs using original documentation
- ## Changes in the programs result in changes in the bitstream
- ## File format
  - 32 byte file-header, variable data block
  - fixed start block (256x 0xFF), separator (128x 0xFF),end block (10x 0x00)
- ## File header
  - filename, file-type, length, checksum, start address
- ## Data Block
  - structure dependent on the file type
  - e.g. BASIC program - for every line: line number, line in ASCII (commands encoded) and start address in memory of next line, e.g.

```
1 0 P R I N T     " H A L L O " <EOL>
0A 00      94       20 22 48 45 4C 4C 4F 22 00
```

Migration Tool

- ## Features

  - Tool to migrate data from audio-streams to non-obsolete formats

  - Either from audio-file or directly from original system

  - Support for all file-types of G7400

  - Data can be encoded in bitstream and/or audio-stream

- ## Two methods for extracting data from waveform

  - existing method from retro community works for data transfer from and to original system but unusable for tapes (too much noise, lost data)

  - novel method that interprets the arc length of the curve instead of counting highs and lows in the signal

Evaluation of Extraction Methods

- **Tests with new files**
  - readable by original system and both methods

- **Tests with 20 years old tapes**
  - no files recognized with original method 1
  - 6 / 23 files recognized by original system (0 without errors)
  - 22/23 files recognized by method 2 (3 without errors)
  - all files recovered from the tapes are BASIC programs

Evaluation Results

- **Validity of files**
    - original system: files loaded were unusable
    - checked by reencoding and testing them on original system (original data for comparison not available)
    - most files had small errors that could be manually restored

- **Data is no longer readable on original system**
- **Evaluated tapes successfully migrated using the migration tool with the method of analyzing arc length of signal**

Conclusions

- ## Data successfully preserved
  - some data lost due to age of tapes (20 years is expected lifetime)
  - media refresh for original system possible by decoding / encoding

- ## Reengineering easier if done now
- test programs can be written and analyzed with access to original system
- expert knowledge still available in "retro" communities

- ## Using bitstream for emulation
- converting only the waveform to bitstream allows use of data in future emulators of the system
- no emulation of C7420 available yet

- ## Results interpreted for other systems/media types
- findings valid for all systems that encode data in audio waveforms
- in most other cases special devices necessary to read media
- logical reengineering and extracting data from bitstreams is possible

Tool and sample files can be downloaded from:

http://www.ifs.tuwien.ac.at/dp/hc_audio_migration

Thank you for your attention.

guttenbrunner@ifs.tuwien.ac.at
www.ifs.tuwien.ac.at/dp
www.planets-project.eu