RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 20 (1995)
Indiana University

Using CD-ROM as a Storage Medium for Digitized Speech Materials[1]

Jon M. D'Haenens and Luis R. Hernández S.

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1]This work supported, in part, by grants to Indiana University–Bloomington: NIH-NIDCD Research Grant DC00111 and NIH-NIDCD Training Grant DC00012.

Using CD-ROM as a Storage Medium for Digitized Speech Materials

Abstract. This report describes the use of CD-ROM technology for the storage of digitized speech material. Advantages of this technology are discussed, as well as general procedures and particular hardware and software specifications. Finally, examples of specific databases from the Speech Research Laboratory that have been stored to CD-ROM are discussed.

Overview

The options available for the storage of digitized audio files used in the acoustic analysis of speech and as stimuli in perception experiments in the Speech Research Laboratory have been limited to magnetic tape storage. Due to the relatively large size of digitized audio files and the increasing number of files that must be stored, it has become impractical to permanently store speech audio materials on the hard drives of computers in our laboratory. As a result of these space and operational limitations, it is necessary to store the materials on high-capacity magnetic media such as TK-70 tape or DAT (Digital Audio Tape). Although these materials have a high capacity and are efficient for off-line storage, they exist mainly for the purpose of archiving data. Sound files must be retrieved from the tape using a utility program and stored on a hard drive before they can be analyzed or used in perception experiments. The result is that researchers using a large volume of digitized speech materials must deal with space restrictions and long waits for the retrieval of sound files from archive tapes.

With CD-ROM technology (re: ANSI, 1990), it is now possible to store a large amount of data on a small disk that provides the advantages of both the tape archival systems and hard disk storage. Some of these advantages include high capacity, speed, permanence, and the ability to work with files directly from the disk. Because CD-ROM is now the standard for many storage applications, the format is supported by nearly every platform, and CD-ROM readers are supplied with almost every new computer. This simplifies the process of sharing data; digitized sound files stored on a disk can be read by any computer with a CD-ROM reader. The data stored on a CD-ROM cannot be erased, written over, or modified, which may be considered an advantage or disadvantage, depending on the application.

CD-ROM is the ideal medium to solve many of the Speech Research Laboratory's digitized speech storage problems. Each disk allows for the storage of a significant amount of digitized audio: approximately 300 minutes of one channel of raw sound data at a 20 KHz sampling rate and with 16-bit resolution. Access to the data is considerably faster than with tape (the access speed is dependent on the CD-ROM reader). Additionally, it is not necessary to copy the sound files to a hard drive to listen to them or access them from within signal analysis software, eliminating the necessity of shuffling files from one storage medium to another. Other advantages include permanence and safety of data storage (CD-ROMs cannot be erased or written over), and the ability to store other relevant textual data along with the audio signals, eliminating most paper documentation and references to the material.

Process

The hardware used for CD-R (CD Recordable) writing in our lab consists of a Macintosh Quadra 840AV with a Yamaha CDE100 CD-R mastering drive connected through the SCSI bus. The Quadra is

connected to the local area network, allowing files to be transferred from almost every computer within the lab. The software used is GEAR for Macintosh®, Version 2.5 (Elektroson Software).

The first step in the creation of a CD-ROM is the restoration of the sound files to the local hard drive of the native platform, for example, from the TK-70 tape to the VAX or from DAT to a local PC. After the files have been restored, they can be transferred directly over the network by copying them to the Quadra equipped with the CD-R mastering drive. An alternative method used for computers with no direct network drive access is transferring the files via FTP (Internet File Transfer Protocol). Usually, the restoration and transfer must be done in batches, as there is typically not enough space to store all files from an archive on a local hard drive. It is important that all files are transferred to the hard drive of the Quadra used to create the CD-ROM, as network access is too slow for the data transfer to the CD.

Once the files have been restored and transferred to the Quadra used for CD-R writing, the GEAR CD-R mastering software is used to create a physical image of the CD. Before creating the image, it is important to choose the ISO-9660 format for the target CD-ROM, which is currently the standard format and is readable by almost every computing platform.

A physical image is simply a bit-for-bit representation of what the CD will contain. Creating a physical image is a necessary step because it saves critical processing and disk access time while the CD is being written. If there is any delay in the data stream during the CD-R writing process, the resulting CD-ROM will be unreadable. After the image has been created, the final step is to place a CD-R in the CD-R mastering drive and select 'Write' from the GEAR menu. At this point, the physical image is written to the disk . After the process is completed, the CD is automatically ejected and is tested to ensure that it was written correctly. For detailed instructions on the use of the GEAR software, refer to the GEAR User's Manual.

Example Applications

In the Speech Research Laboratory, we have already created two CD-ROMs, each with a slightly different purpose. The PB/MRT database (re: ANSI, 1971; House et al., 1965) of monosyllabic words is used very often by many people in the laboratory, and is often requested by people outside the laboratory. In the past, these data were either stored on a hard drive so other labs could access them over the Internet, or the desired parts would be retrieved from tape and stored on a hard drive for individual use. Putting these data on a CD-ROM makes them much easier to share. The CD-ROM can be mounted on a PC and accessed via FTP from other laboratories, saving retrieval time and disk space. Additionally, whenever anyone needs to access the data within the laboratory, they can simply insert the CD-ROM into any computer with a CD-ROM reader and have immediate access to the data on the disk .

Another CD-ROM that we created was an archive of materials obtained for the Indiana University/General Motors Research Laboratories studies of the effects of alcohol on speech. This archive contains digitized sound files of words, sentences, and paragraphs. The primary motivation for putting these data on CD-ROM was the anticipation of doing a large amount of computer aided acoustic analysis of the files. By storing these data on CD-ROM, not only do we save time and disk space, but our researchers have immediate and random access to all of the materials, making the analysis much faster and easier.

For the GM studies CD-ROM, the data was initially written to the CD-ROM in the ".ils" format (Interactive Laboratory Systems: Signal Technology, Inc.) in which that data were digitized. Because the

analysis was to be performed with software without native support for the ".ils" format, the files were transformed to the Entropic Waves ".sd" format (Entropic Research Laboratory, Inc.) and written to another CD-ROM. This was necessary to eliminate the need to convert the files from one format to another and the need to store them on a hard drive.

Summary

In the process of creating the two CD-ROMs described above, we have arrived at some important points to consider before beginning the process.

- Because a CD-ROM is a write-once medium, it is important to make sure everything to be placed on the CD is in its final form and has been verified. If anything needs to be changed after the CD-ROM has been written, a new CD must be burned in (created).

- Although it is possible to write the data in multiple sessions, some CD-ROM readers will not recognize any session after the first. If planning on writing multi-session CD-ROMs, verify that the CD-ROMs drives to be used are multi-session.

- Before beginning the process, decide on a final directory structure for the organization of the CD-ROM. This makes the process of transfering the files easier, and a good plan makes working with the CD-ROM much easier.

- Use the ISO-9660 format. It is the most universally recognized format.

Conclusions

Although we are in the early stages of utilizing CD-ROM technology, it has already proven to be reliable and time-saving. CD-ROM is not perfect for every storage application because of the overhead in creating the disk , but there are many ways that the Speech Research Laboratory has found it to be useful for archiving large sets of sound files. The next steps are to create CD-ROMs for the remainder of the databases that are often used, and to educate the users of files in our lab on ways it can be utilized.

References

American National Standards Institute. (1971). *Method for measurement of monosyllabic word intelligibility* (American National Standard S3.2-1960 [R1971]). New York: Author.

American National Standards Institute. (1990). *Volume and file structure of CD-ROM for information interchange* (ANSI/NISO/ISO 9660-1990). New York: Author.

House, A.S., Williams, C.E., Hecker, M.H.L., & Kryter, K.D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America*, **37**, 158-166.

Pisoni, D.B., & Martin, C.S. (1989). Effects of alcohol on the acoustic-phonetic properties of speech: Perceptual and acoustic analysis. *Alcoholism: Clinical and Experimental Research*, **13**, 577-587.

Pisoni, D.B., Yuchtman, M., & Hathaway, S.N. (1986). Effects of alcohol on the acoustic-phonetic properties of speech. In *Alcohol, Accidents, and Injuries* (pp. 131-150). Warrendale, PA: Society of Automotive Engineers.