Managing Stored Voice in the Etherphone System

Douglas B. Terry
Daniel C. Swinehart
Computer Science Laboratory
Xerox Palo Alto Research Center

Extended Abstract

The Etherphone^{TM} system was developed at Xerox PARC to explore methods of integrating voice into existing distributed personal computing environments. An important component of the Etherphone system, the *voice manager*, provides operations for recording, playing, editing, and otherwise manipulating digitized voice based on an abstraction that we call *voice ropes*. It was designed to allow:

- unrestricted use of voice in client applications,
- sharing among various clients,
- editing of voice by programs,
- integration of diverse workstations into the system,
- security at least as good as that of conventional file servers, and
- automatic reclamation of the storage occupied by unneeded voice.

As with text, we want the ability to incorporate voice easily into electronic mail messages, voice-annotated documents, user interfaces, and other interactive applications. Because the characteristics of voice differ greatly from those of text, special mechanisms are required for managing and sharing stored voice. The voice manager reduces the work generally associated with building voice applications by providing a convenient set of application-independent abstractions for stored voice.

Clients view *voice ropes* as immutable sequences of voice samples referenced by unique identifiers. In actuality, a voice rope consists of a list of intervals within *voice files* that are stored on a special voice file server. A database stores the many-to-many relationships that exist between voice ropes and files. Maintaining voice on a publicly accessible server facilitates sharing among various clients. Clients can freely share references to voice ropes without incurring the overhead of transmitting the voice itself. To ensure privacy, access control lists govern who is permitted to play or edit particular voice ropes.

Rather than rearranging the contents of voice files to edit voice, the voice manager simply creates new voice ropes from old ones and adds them to the database. Figure 1 depicts three voice ropes consisting of intervals from two voice files; the third voice rope could have been created by combining the first two using the voice rope editing facilities. The editing operations provided by the voice manager closely resemble operations normally associated with text string manipulation. This is intentional so that programmers can manipulate voice in the ways to which they are accustomed to dealing with text. The basic facilities to support editing reside on a server; workstations are responsible for providing a user interface that is integrated with their programming environment.

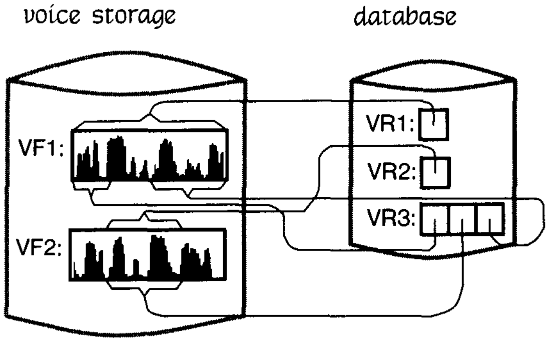

Figure 1. Two level storage hierarchy: voice ropes refer to intervals of voice files.

Clients register *interests* in particular voice ropes. Interests, additional persistent data structures maintained by the server, serve two purposes. First, they provide a sort of directory service for managing the voice ropes that have been created. More importantly, they provide a reliable reference counting mechanism, permitting the garbage collection of voice ropes that are no longer needed. These interests are grouped into classes; for some important classes, obsolete interests can be detected and deleted by a class-specific algorithm that runs periodically. Devising techniques for automatically collecting garbage in a distributed, heterogeneous environment was one of the most difficult problems faced in the design of the voice manager.

These facilities for managing stored voice in the Etherphone system were designed with the intent of moving voice data as little as possible. Once recorded in the voice file server, voice is never copied until a workstation sends a play request; at this point the voice is transmitted directly to an Etherphone, a microprocessor-based telephone instrument. In particular, although workstations initiate most of the operations in the Etherphone system, there is little reason for them to receive the actual voice data since they have no way of playing it.

Adding such voice facilities to a diverse and complex software base presents challenging problems to the systems builder since much of the existing workstation and server software cannot be changed or extended. Manipulating stored voice solely by textual references, besides allowing efficient sharing and resource management, has made it easy to integrate voice into documents. The only requirements placed on a workstation in order to make use of the voice services are that it have an associated Etherphone and an RPC implementation.

The Etherphone system uses secure RPC for all control functions and DES encryption for transmitted voice. These ensure the privacy of voice communication, which is important even in a research environment, although the network is inherently vulnerable to interception of information. Storing the voice in its encrypted form protects the voice on the server and also means that the voice need not be reencrypted when played. All in all, the voice manager provides better security than most conventional file servers.

The performance of operations for editing and managing recorded voice must be compatible with human response times: sub-second response at a peak rate of several operations per second is more than adequate. Performance measurements confirm that the voice manager easily meets these requirements.

In conclusion, the major technical contributions presented in this paper involve the use of simple databases to:

(1) describe the results of editing operations such that existing voice passages need not be moved, copied, or decrypted, and

(2) provide a modified style of reference counting that allows the automatic reclamation of obsolete voice.

Approximately 50 Etherphones are in daily use in the Computer Science Laboratory at Xerox PARC. We have had a voice mail system running since 1984 and a prototype voice editor available for demonstrations and experimental use since the spring of 1986.