ELECTRICAL SCIENCE

**WHITE PAPER:**

# TECHNICAL DESCRIPTION OF VOICE LOGGER RETRIEVAL TECHNIQUES

**Andrew Stevens, Ph.D.**

**ABSTRACT—Many businesses and agencies use voice loggers to record telephone traffic, but retrieval of the archived recordings can be difficult, especially in bulk. This article addresses many of the issues related to exporting audio and metadata from voice loggers so they may be searched, replayed, and analyzed outside of the original environment. Topics include: estimation of the amount of data, conversion of archive media to standard file formats, interpretation of per-call metadata, requirements for repositories for the retrieved audio and metadata, and adherence to compliance rules. Applications include e-discovery for litigations, migrations to new logger platforms, and ingestion into voice analytical engines.**

---

Voice logging is the act of recording telephone and other audio traffic for the purpose of creating unambiguous records of conversations and their specific dates and times. It is often used for regulatory compliance, quality assurance, or public safety purposes. Typical applications include financial trading floors (to verify buy and sell orders or improper trading), telephone call centers (to study agent effectiveness and customer satisfaction), and emergency hotlines (to monitor response times or provide evidence for criminal trials).

Voice loggers are the physical equipment used to create and store the recordings as well as their associated timestamps, phone numbers, call durations, and other related fields (collectively referred to as the “per-call metadata”). Voice loggers may be connected to individual telephone extensions, open microphones (“hoot-n-holler boxes”) or trunk lines, and they record all calls and metadata as they occur. Depending on the particular installation, voice loggers may need to digitize audio from analog phone lines; in other cases, the audio is already digital (e.g. for T-1 trunks or VoIP). Voice logger equipment is designed to support a specified number of “channels” corresponding to the maximum number of calls which may be recorded simultaneously (typically 32 or more). A voice logger must also have an accurate clock (to record timestamps) and often has a number of customizable fields such as agent name, trunk name, or case number. A voice logger may also have special circuitry to detect dialed digits within a phone call (DTMF tones) or minimum levels of activity (so that it can stop recording during long periods of silence).

In addition to the recording of voice traffic, all voice loggers also provide the capability to replay calls based on date or channel and can export retrieved audio into a standard format such as WAV (for audio) or CSV (for metadata). Typically, a given logger is capable of exporting calls in small quantities (perhaps a few dozen at a time) and, especially for older loggers, require that the


Fig. 1. Common voice logger archive media: DVD-RAM, AIT-1, DDS-3

appropriate archive tape or disk to be inserted at the front panel of the machine. Thus, retrieving even a few hundred calls can be a slow and arduous process through the logger's built-in interface. And while this native retrieval capability suffices in some instances, there are many scenarios where it is sadly lacking. For example, many companies decide to change logger platforms and then realize that they need to retain their old equipment in order to replay their old media, while also finding that the jilted former provider is reluctant to offer support. In other cases, the logger equipment manufacturer has been acquired and their respective product line subsequently discontinued by the acquirer (e.g. Racal and Dictaphone acquired by NICE, or Eyretel acquired by Verint). In these cases, there is no official support and logger owners must find the few remaining resellers who still have parts, or perhaps resort to buying used equipment on eBay.

There are many other scenarios as well, and this white paper is intended to address the challenges of voice logger audio retrieval as faced by many logger owners who need to produce audio from their logger archives due to pending litigation, compliance regulations, or hardware migration. It will also describe the advanced techniques developed by Electrical Science which allow bulk retrieval beyond the limited capabilities of the original loggers.

**BACKGROUND**

Because of the large amount of generated data from recordings and metadata, voice loggers archive the audio to a variety of mass storage devices. Older units use physical media such as magnetic tapes or optical disks (Fig. 1). Newer recorders may archive recordings and metadata to hard disks, network file shares, network-attached storage (NAS), or enterprise data archives (e.g. EMC Centera). In most cases, the tapes, disks or files are fundamentally different from conventional hi-fi audio recording media (e.g. cassette, CD, MP3) and are designed to be replayed only from the original equipment. Each logger manufacturer uses a different binary format for writing the actual bytes onto their tapes or archives, and for example, a DDS-3 tape written on one logger (e.g. Racal Wordnet Series 2) will not play back on any other logger model, even if that logger also reads and writes DDS-3 tapes (e.g. NICE 8.9 or Eyretel MediaStore).

Like any computing device, voice loggers require maintenance and become obsolete over time, with a typical lifetime of 5-10 years. However, compliance regulations and lengthy litigations often dictate that the recorded audio be retrievable for longer intervals, thus requiring that the old equipment be kept working. For example, an insider trading inquiry might take years to come to light and might subsequently involve more years of discovery and legal proceedings. A typical investigation might necessitate the production of all recordings for a given stock trader for an interval of months or years. The corresponding recordings (often numbering in excess of 10,000 recordings for each trader) need to be retrieved quickly and accurately, must be of a format acceptable to all parties and the courts, and must be technically defensible in a court of law.

Unfortunately, most voice loggers are not designed to perform bulk retrievals. For example, the NICE 8.9 recorder limits retrievals to 100 calls at a time. Thus, a retrieval of 10,000 calls is a laborious process when performed on the original recorder. Alternatively, in other (non-legal) contexts, businesses which wish to migrate to a new logger platform might have millions of calls trapped on tapes or disks generated by the old logger hardware. Yet most logger manufacturers provide no way to export these millions of stored calls outside of the original equipment, often for the express reason of discouraging customers from switching vendors.

## HOW MUCH DATA?

A common mystery regarding voice loggers is the amount of audio stored in each disk, tape or file. For example, for a single AIT-1 tape (35GB capacity) of stored recordings, how many hours of audio are present, how many individual recordings are present, and how much storage will be necessary to store the retrieved audio files? Or for 10 terabytes of logger files on a NAS, how much audio is there? The answer is not straightforward and depends on the logger equipment design, including the method used to compress the audio (the “codec”) and the possible suppression of silence when storing the recordings.

All digital recording systems employ an audio codec for converting analog signals (as captured by a microphone) into bytes. In the case of speech recording, the codec is optimized to minimize the bandwidth and storage requirements while maintaining (or even enhancing) voice intelligibility. Note that these requirements are vastly different than those for hi-fi audio reproduction, where accuracy and linearity are paramount to the extent that understanding the speaker’s words is of secondary concern.

Technically speaking, a speech codec is a mathematical algorithm which maximally compresses the audio while retaining intelligibility and dynamic range. Early codec algorithms (e.g. G.711) were necessarily simple due to the minimal computing power available in the 1970s. As microprocessors have improved, so have codecs, and modern codecs (e.g. G.729 or G.723.1) compress around ten times as well as their older counterparts.

The storage efficiency for a given codec is indicated by its stated bitrate which corresponds to the amount of disk space necessary to store one second of audio. For example, G.711 encodes audio at 64000 bits per second (or bits/sec or “bps”) and so each recorded second requires 64000 bits (or 8000 bytes) for storage, and a 1-hour recording requires 28.8 megabytes of space. For the more modern G.729 codec, each recorded second consumes 8000 bits (or 1000 bytes), an 8X improvement over G.711, and thus the same 1-hour recording consumes only 3.6 megabytes. On the other hand, G.729 requires a much more powerful CPU than G.711 in order to implement its complex mathematical algorithm. In addition, its playback quality is not quite as good as G.711 because its compression algorithm can slightly distort the audio upon playback (though in fact, G.711 and G.729 are largely indistinguishable to ordinary human ears). Some common voice logger codecs and corresponding bitrates are listed in Table 1.

Another factor which can affect stored data size is that many loggers employ silence suppression to avoid wasting disk space during lengthy silences, e.g. if a call gets placed on hold. For

**Table 1. Common voice logger codecs and bitrates**

| CODEC | BITRATE (bits/sec) |
|---|---|
| G.711 (μ-law, A-law) | 64000 |
| Vox (Dialogic ADPCM) | 24000 |
| G.726 (CCITT ADPCM) | 16000 |
| GSM 06.10 | 13000 |
| G.729 | 8000 |
| TrueSpeech | 8000 |
| G.723.1 | 5300 |

**Table 2. Common voice logger media types and capacities**

| MEDIA TYPE | CAPACITY (GB) |
|---|---|
| DDS-2 | 4 |
| DDS-3 | 12 |
| DDS-4 | 20 |
| AIT-1 | 25, 35 |
| AIT-1 Turbo | 40 |
| AIT-2 | 36, 50 |
| DVD-RAM | 4.7 per side |

example, if a call uses the G.729 codec and contains 30 seconds of silence, the logger might store a brief message indicating the silence duration (perhaps consisting of only a few bytes) instead of writing the entire 30000 bytes required to store the 30 seconds of silence.

Thus, assuming that we know the codec and also the nature of the recorded audio (and how much silence it contains), only then can we accurately predict how many hours of audio are contained on a 35GB tape or 10TB NAS. Unfortunately, oftentimes neither item is known. However, based on prior experience, a very rough estimate for a modern logger might assume an 8000 bps bitrate (corresponding to G.729 codec and minimal silence suppression), resulting in 9722 hours of converted audio (35,000,000,000 bytes × 8 bits/byte ÷ 8000 bits/sec ÷ 3600 sec/hour) which could fit on a single 35GB tape. Then again, if the logger was configured for G.711 encoding (at 64000 bits/sec), the amount would be the much smaller maximum of 1215 hours. A similar calculation for a 10TB NAS implies that it could hold approximately 2.8 million recorded hours (for G.729 encoding) or 350,000 hours (for G.711).

Note that for magnetic media, most tape manufacturers quote both a “native” and a “compressed” capacity. Because voice logger audio is already compressed by the codec, it will not get additionally compressed by the tape drive and thus only the native capacity should be used in any calculation. Typical voice logger media types and corresponding native capacities are listed in Table 2. Most logger manufacturers use off-the-shelf drives from Sony, HP, and Matsushita. Dual-spool helical-scan magnetic tapes are popular due to their low cost and backward compatibility (e.g. DDS-4 drives will read the older DDS-3 and DDS-2 tapes). Optical media (e.g. DVD-RAMs or MO disks) are more reliable than tape but have lower capacity and are usually used in smaller installations.

As an example, the Racal Wordnet Series 2 recorder was designed circa 1999 and used DDS-3 tapes (12GB capacity) and the GSM 06.10 codec (13000 bps), both of which were state-of-the-art at the time. This storage capacity and bitrate implies that a maximum of 2051 hours of audio could fit on a single tape (12,000,000,000 bytes × 8 bits/byte ÷ 13000 bits/sec ÷ 3600 sec/hour). The actual amount is somewhat less due to overhead for indexing and metadata. The official tape capacity for Wordnet Series 2 as specified by the original Racal literature is 1980 hours per tape.

## TARGET AUDIO FORMATS

When delivering extracted audio from voice loggers, the choice of delivery format depends on the target audience. The most common target is playback for human ears via the default audio player under Windows, MacOS, iOS, or Android. In this case, WAV format is a good choice because it is supported in all environments. WAV is a “container” format, meaning that it can be used to package audio which is encoded with any specified codec. However, just because an individual WAV file uses a given codec does not automatically mean that codec’s playback algorithm is available in the playback environment. In fact, many common speech codecs are not preinstalled by many operating systems, e.g. G.729 is not delivered by default with Windows (though it can be installed separately). In order to avoid a software installation, we have found that the codec which has the broadest availability amongst today’s popular playback platforms is GSM 06.10 (13000 bps), making it the optimal choice in most cases. (Note that while the MP3 codec is also available on most platforms, this codec is intended for high-fidelity music and is not well-suited for low-fidelity speech which is bandlimited in telephony to 3 kHz.)

In some cases, the target for the retrieved recordings is not the human ear but rather a speech analytical engine. Various companies specialize in audio software to enable keyword searching, voice identification, customer satisfaction metrics, and other analyses on recorded speech. Because these engines perform extensive mathematical analysis on the raw audio bytes, the preferred codec in this case is G.711 in order to obtain the best possible interpretation of the digitized audio and also minimize the number of codec conversions (each of which can slightly degrade the recording). Note that converting to any codec with higher quality than G.711 (e.g. 16-bit linear PCM) serves no purpose because most modern voice traffic inevitably gets converted to G.711 (or higher compression) at some point during its end-to-end journey between calling parties. (Analogously, once a high-resolution digital photo has been reduced to low resolution, the photo cannot be converted back to high resolution.)

## METADATA

All voice loggers store metadata related to each recording, always minimally including the call start time, stop time, and channel number. Depending on how the recorder was originally installed, the channel number might indicate the desk extension of the user (if installed on the “station” side of the PBX) or a trunk number (if installed on the network or “PSTN” side of the PBX). Most loggers also include a configurable per-channel text field in order to store a useful label indicating the source of the audio (e.g. “x6596” or “John Smith”). It is up to the logger operator to maintain the accuracy of these fields. Alternatively, some logger operators may choose a different method for storing the channel information, sometimes resorting to a spiral notebook for manually handwriting which agent is present on which channel on each day (or week or year).

More sophisticated loggers will store “extended” metadata which might also include agent name, agent ID, extension number, extension type (e.g. handset or microphone), trunk name, trunk ID, incoming/outgoing flag, outgoing dialed digits, incoming dialed digits (DNIS), detected DTMF tones, CLI (caller line identification, aka caller ID), and more. Some loggers also support installation-specific custom fields which might contain business-related information like invoice number or customer name. The availability of these fields depends on the capabilities of the logger hardware, its ability to connect to the PBX via CTI (computer-telephony integration), integration with CRM (customer relationship management) software, and the proper setup and configuration of the integrations. In some logger models, the extended metadata capability is optional (e.g. NICE 8.9 with optional CLS database), thus allowing smaller installations to save money (e.g. emergency hotlines at a local police station). In other cases, the extended metadata is critical (e.g. at contact centers or financial trading floors).

Depending on the logger hardware, the association of the metadata to actual recordings might be "implicit" (because they are stored together on the same magnetic tape) or might require a matching operation (because the extended metadata are stored separately from the audio). Examples of loggers which archive the metadata together with the audio are Racal Wordnet, Racal Mirra, Eyretel, Verint Impact, and Witness. Examples of loggers which store the extended metadata separately from the audio include NICE 8.9 with CLS, NICE Perform, CyberTech, NICE NTR, and Dictaphone Freedom. While storing the metadata in a separate database can be convenient (resulting in a more modular system architecture and enabling more sophisticated reporting), it can also lead to problems for retrieval if both subsystems are not backed up properly and synchronized. We have seen some cases where the database has failed independently of the recorder, resulting in thousands (or even millions) of recordings with missing metadata because nobody noticed the database failure. In other cases, the logger is retired and the archived audio is safely stored, but nobody remembers to backup the database.

When extended metadata is stored separately from the audio, the mechanism for matching the two is usually straightforward, but interpretation of the combined metadata sets can be confusing. This is because the method for determining the beginning and end of an individual call can vary depending on the particular logger hardware, the PBX model, and the method of their interconnection. Generally speaking, a telephone "call" is considered to begin when the calling party picks up their handset (or colloquially speaking, "goes offhook"), and ends when one of the parties hangs up (or "goes onhook"). The "call detail record" (CDR) for each individual call is stored in a database inside the PBX. Because the offhook/onhook events are only detectable by the PBX, the corresponding CDRs are only available to the voice logger if it is integrated with the PBX via a CTI connection.

In cases where the logger is *not* integrated with the PBX, then the only way the logger can detect offhook/onhook is by watching for activity on the monitored voice circuits, also called "VOX" detection. But a problem can arise when a call becomes too quiet (e.g. if the call is placed on hold) and the logger is fooled into thinking that the call has ended and stops recording. When the call resumes, the logger starts recording again, but has no way of knowing if the new activity is a new call or resumption of the previous call. Thus, a single CDR "call" (as determined by the PBX) might consist of several VOX "recordings" (as determined by the voice logger). Stated another way, CDR calls and VOX recordings can have a "one-to-many" relationship and require expertise and experience to weave together into a single contiguous unit.

In another scenario, a caller might hang up a call and then immediately place a new call, with only a short period of silence between calls. In this case, the logger does not always detect the hangup and will combine the two calls (as determined by the PBX) into a single recording (as determined by the voice logger). Thus, two CDR calls in the PBX will correspond to a single VOX recording in the logger, also called a "many-to-one" relationship.

In another scenario, many trading floors employ open microphones ("hoot-n-holler" boxes) which are connected to the PBX and record all activity, 24 hours per day. From the PBX's perspective, the open microphones are treated identically to ordinary handset calls, but can have durations of months (or even years) per call. On the other hand, voice loggers usually have a maximum recording length, typically in the vicinity of three to twelve hours. Thus, the CDR for a single open-microphone call in the PBX (perhaps having a duration of six months) can correspond to hundreds of recordings in the voice logger. Further complicating the scenario is that trading turrets often have multiple open microphones (sometimes as many as ten per turret) which are multiplexed on to a single logger channel in order to reduce cost (because voice loggers are often priced per channel). Thus, for open microphones, a “many-to-many” relationship can exist between CDR calls and VOX recordings.

The upshot is that some voice logger installations have two distinct sets of metadata: (a) the VOX metadata corresponding to the actual recordings (as determined by voice activity), and (b) the CDR metadata corresponding to the actual calls (as determined by offhook/onhook). And whereas the original logger playback environment has the ability to reassemble on-the-fly the VOX recordings into CDR calls (by truncating or padding the recordings with silence as necessary), the target repository for the exported audio and metadata often does not have this capability. Consequently, confusion can arise during the review of the exported recordings due to the apparent “mismatch” of the CDR metadata and their component VOX recordings. The solutions to the confusion are to (a) educate the reviewers about the intricacies of voice logger metadata, or (b) perform a wholesale reconstruction of the CDR calls from their VOX recording components, with the penalty of requiring additional disk space to store the reconstructed calls due to reinserted silence and channel demultiplexing (which, based on previous experience, can cause a 100X increase in the required disk space to store the resulting WAV files).

## TARGET AUDIO REPOSITORY

Depending on the size of a given retrieval project, there are many available delivery options for the produced metadata and WAV files for subsequent search and replay. For smaller jobs (e.g. a single tape or fewer than 100,000 recordings), Microsoft Excel is a simple and familiar format for the metadata. In this case, the spreadsheet can be easily searched for interesting calls (based on date, channel, etc.) and then the corresponding WAV files can be replayed via hyperlinks in the spreadsheet. The entire retrieval (spreadsheet plus audio) can usually be written to a DVD-ROM or flash drive.

For medium-size projects (e.g. a few dozen media or fewer than 2 million recordings), metadata delivery in a Microsoft Access database is a good option. Access can handle several million records with good performance and is easily customizable to support most metadata sets. Software licenses for Access are readily available and inexpensive. The metadata is searched by entering criteria into an Access form which then generates tabular search results with embedded hyperlinks to the audio. For delivery to the customer, we store the Access database and WAV files onto a portable USB external hard disk. The customer can then search and replay the audio directly from the hard disk, or alternatively copy the database and audio to their corporate file server to provide access by personnel across the company.

For larger projects (more than 50 media or greater than 2 million recordings), several options exist for delivery with varying levels of cost and complexity. Some companies have existing digital asset repositories (either hosted in-house or at a vendor), and metadata and audio delivery can be customized to meet the import requirements of the repository (typically XML or CSV format for metadata and WAV for audio). In other cases, the customer is migrating logger platforms and would like the legacy audio to be imported into their new logger environment. This is achievable with varying levels of difficulty ranging from extremely simple (i.e. with full cooperation by the new logger vendor) to extremely complex (i.e. without any vendor cooperation and with the added risk of voiding the warranty on the new logger). A third option is the installation of a new audio repository which is specifically tailored for voice logger metadata search and replay. We have worked with several partners to deliver such repositories which consist of a relational database and a Web-based server application to provide browser-based search and retrieval capabilities. Advantages include the ability to support multiple simultaneous users and the avoidance of any application-specific software installation on the client device.

## RETRIEVAL METHODOLOGY

There are several approaches to retrieving voice logger audio, including: (a) using the original logger equipment, (b) using a bulk retrieval tool supplied by the original manufacturer (assuming such a tool exists), (c) writing a custom tool to query the original equipment via an existing API (which may or may not be documented or approved by the manufacturer), or (d) extracting the audio and metadata directly from the archive media or files, without use of any of the original logger hardware or software.

As mentioned earlier, retrieval via the original logger, whether via the front panel or by using a GUI application, is slow and inefficient. Some older loggers are designed to only allow one-at-a-time playback through a speaker on the front panel, while some newer models allow retrieval of up to 100 recordings at a time via a GUI application. In the latter case, it is possible to employ multiple operators at multiple workstations in order to extract thousands of calls per day. However, this technique is vulnerable to quality problems due to its dependency on human keystrokes and mouse clicks in order to save each group of 100 retrieved calls.

Some manufacturers provide “playback only” environments (e.g. Eyretel eWare) or “extraction toolkits” (e.g. NICE ETK). However, these solutions are often expensive and slow because they still depend on logger equipment. For example, if the installed logger has only one tape drive, then retrieval can only proceed one tape at a time.

In cases where the logger supports client-server communication over a network (e.g. NICE 8.9 or Racal Wordnet), it is possible in theory to write a program which emulates the client software and performs repeated retrieval requests without human intervention. However, such a program is still limited by the retrieval speed of the original equipment.

Having attempted all of these approaches, we have determined that the fastest and most reliable approach for bulk retrieval is to extract the audio files directly from the archive media, without use of the original equipment and/or software. While the difficulty of this task varies from logger manufacturer to manufacturer, the advantages in accuracy and speed can be enormous. In cases where the archives are tape or optical disk, the speed improvement can be a factor of 10X or more. And by using generic tape/disk readers and servers, the retrieval process can be scaled up by simply adding more readers and servers to the job.

Retrievals vary from logger to logger but generally follow this procedure:

Step 1: **Copy** all archived data to scratch space, typically a hard disk. This involves byte-for-byte imaging of magnetic tapes and optical disks, file copy from disk partitions and network-attached storage (NAS), bulk retrieval from enterprise storage (e.g. EMC Centera), and table export from relational databases (e.g. Microsoft SQL Server or MySQL). Copying all of the media up front allows immediate detection of any bad, damaged, or unexpected blank tapes and disks which may need data recovery. Note: we typically see approximately 5% of tapes on a given project with some kind of read error, but we have developed techniques to recover data from damaged tapes with high success rates. Optical disks and magnetic storage are generally much more reliable.

Step 2: **Extract implicit metadata** from the raw data and reconstruct into individual call records, one record per recording. Depending on the logger, the available fields may be minimal (limited to call start/stop times and channel number), or may be rich (including agent name, dialed digits, caller ID). For focused retrievals (e.g. limited to specific dates or channels), use the metadata to determine which tapes, disks, or files should undergo audio retrieval.

Step 3: **Extract audio** bytes from raw tape images, raw disk images, and/or audio files. Demultiplex interleaved audio and reinsert suppressed silence as necessary to reconstruct into individual recordings.

Step 4: **Transcode** the audio because loggers often use high-compression codecs which are not available on common audio playback software on Windows, Mac, iOS, or Android. Reformat the audio to a standard format, typically: WAV file with GSM 06.10 codec.

Step 5: **Extract extended metadata** from relational databases (if present). Because some loggers store call detail records (CDRs) separately from the archived audio, this extra step is necessary to extract the CDR call metadata and match it to the VOX recording metadata.

Step 6: **Deliver** the retrieved audio and metadata to an appropriate destination location (typically a portable hard drive) for subsequent import into an audio repository, e-discovery review platform, voice analytics system, or modern voice logger system. Requirements might include file naming conventions (typically WAV files named by channel and date), directory hierarchies, and metadata format (typically CSV or XML).

Note that the above procedure is adjusted depending on requirements of the individual retrieval. For example, in a migration from old logger hardware to new, the retrieval might be a one-time bulk conversion for millions of old recordings and might take weeks or months to perform (depending on the amount of legacy data) and require multiple servers. In addition, many of the steps above can occur in parallel.

Alternatively, live sites may only need to retrieve each day’s recordings (perhaps a few thousand) for subsequent ingestion into a voice analytics engine; in this case, a small networked server appliance (installed locally or cloud-based) can perform all of the steps on a daily (or even hourly) basis, including fetching of the recordings from the file system and fetching of metadata via API connection to the live logger database.

## LOGISTICS

Because voice loggers are often installed in order to satisfy regulatory compliance rules, the same rules will often limit the way retrievals can be performed. Compliance agencies and standards such as the SEC (USA), CFTC (USA), FCA (UK), FSC (South Korea), CSRC (China), ASIC (Australia), and PCI DSS (global) have strict rules relating to data security, data retention, and the transportation of backup media. Depending on the nature and jurisdiction of the recordings, sometimes the voice logger recordings must remain on premises; in other cases, they cannot leave the country; in other cases, the original tapes cannot leave the premises but tape copies can be shipped for retrieval; in some cases, the media can only be shipped by “white glove” courier.

As a result, voice logger retrievals must sometimes be performed on premises at the recorder location. A major advantage of our use of generic hardware is that the equipment for retrieval is low cost and readily available, and may even be supplied by the customer. The retrieval can be done in one of several ways, including: (a) remotely (via SSH or VPN) on customer-supplied Linux servers; (b) on a portable retrieval environment (containing rackmount tape readers and servers) which is transported to the premises and operated by trained engineers; or (c) for smaller ongoing retrievals, via a rackmount server appliance which is installed in the customer LAN environment and is integrated with their logger and database for automated daily (or hourly) retrievals.

**SUMMARY**

This white paper has attempted to describe some of the issues related to the bulk retrieval of audio from voice loggers. Some readers may be surprised at the complexity of the task which arises from a combination of the complicated technical architecture of the original logger hardware, the stringent limitations of various governing bodies regarding storage and retention, and the competitive and short-lived nature of voice logger manufacturers (as evidenced by the disappearance of many vendors including CyberTech, Dictaphone, Eyretel, Mercom, and Racal). We have performed hundreds of successful retrieval projects worldwide of hundreds of millions of recordings over many years from most major logger models for financial institutions, contact centers, civil and criminal investigations, and public safety organizations. Electrical Science is the unmatched global expert in voice logger retrieval.

---

**ABOUT THE AUTHOR**

Andrew Stevens has been working with all types of telecom equipment ever since he was a telephone repairman as an undergraduate at MIT. After receiving his bachelor's degree in 1986, he worked at Bell Laboratories and IBM, and in 1995, he received his Ph.D. in electrical engineering from Columbia University. Since 2001, he has been president of Electrical Science, an engineering consulting firm specializing in finding elegant solutions to difficult and esoteric problems. For the past twelve years, Electrical Science has been the leading supplier of voice logger retrieval services for litigation support, trading floors, telephone call centers, and public safety hotlines.

ELECTRICAL SCIENCE

Electrical Science, Inc.
114 Pearl St., Suite 2B
Port Chester, NY 10573 USA
+1-914-939-7396
www.electricalscience.com