



AUDIO: Searching is Different than Documents

By Jeff Schlueter*

You have been made aware of potential litigation, a regulatory inquiry, or an internal investigation. You've instituted a litigation hold, located the responsive electronically stored information (ESI), and now it is time to begin an initial culling of your content. A preliminary analysis reveals that in addition to documents, you have voice mail embedded in your unified messaging system, hundreds of hours of trading floor or call center recordings, and archived web conferences that are potentially responsive. Do you review audio using the same methodologies as your other traditional documents?

The simple answer is "No," but not for the reasons you may think. Humans simply do not speak the same way they write. Even if you choose to transcribe all of your audio content and use traditional eDiscovery review tools, the content must be searched using different techniques.

Three specific elements differentiate the spoken word from writing:

- **Standardization** – While spoken languages have dialects that vary by geography, social and economic factors, written communications are far more structured.
- **Formality** – Speech is typically a more casual form of communication, while written communications are more structured and traditional. This certainly is a function of the permanence of written communication and the thought process involved in creating a written document, as opposed to the impromptu nature of speech. Verbal 'shorthand' is typical, and often used in familiar settings.
- **Change** – Our speech is constantly undergoing changes, often so subtle we are not even aware of them. New words, terms, and phrases typically receive life in speech, and once universally adopted may find their way into written

communications. New words and phrases have become so prevalent that both the Merriam-Webster Dictionary and the Oxford English Dictionary have complete sections of their websites dedicated to new words and their usages.

These elements combine to make traditional document search techniques, even of transcribed recordings, inadequate. Three examples:

- In publicly available recordings from Enron, energy traders discuss a specific California utility, the Imperial Irrigation District. However, a search for the term “Imperial Irrigation District” returns few results in a search of audio content or transcriptions. The reason is simple: instead of saying the whole company name, the traders use the acronym “IID”. However, in the recordings, IID is spoken as “Double I D.” In order for the search to be successful, reviewers need to understand vernacular of the speakers.
- Two sports enthusiasts are engaged in a discussion of college football. One of them mentions the “NCAA.” “Are we searching for “N C A A,” “N C 2 A.” “N C Two A,” or “N C Double A”? In written documents, we would search for only “National Collegiate Athletic Association” and “NCAA.”
- In a criminal investigation, wiretapped telephone recordings revealed that terms related to specific criminal acts had been replaced by common words and phrases. These seemingly innocent conversations took on a completely different tone once the terminology was decoded.

Automating Audio Searches

Clearly, using keywords derived from text search against your audio content are likely to reveal some, but likely not all, of the relevant content. While the keywords you use to locate documents are a good place to start with your audio, nothing replaces the astute, trained ear. As you delve into hundreds or thousands of hours of recordings, the prospect of listening to each recording is not only time and cost prohibitive, but human listening fatigue can easily render the entire effort worthless.

It is estimated that the time required for humans to review audio content for audio discovery purposes is 4:1; that is, it takes four hours of listening for every one hour of recorded content. Considering a volume of as little as 500 hours of audio, the human listening review process would require approximately 2,000 person-hours to complete. Using a team of five reviewers, working full-time on the project, the first review would require ten weeks. Most compa-

nies simply do not have the manpower or resources to throw at a project of this magnitude.

One alternative, transcription, is also fraught with challenges to successful searches. First, similar cost factors are associated with transcription as human listening. At an average of \$150 per hour of content for professional transcription services, the 500 hours of content would cost \$75,000 for the transcription, before any review. Next, the quality of the transcription is only as good as the transcriber, and the transcriber’s level of understanding of the content. Transcribers are trained to transcribe exactly what they hear and not to interpret what they think they hear. Will the transcriber type “Double ID,” “Double I D,” or “IID”?

Like documents, the metadata stored in voicemail and call-logging systems may be extremely valuable in culling the volume of audio content. Call lengths, call extensions, and caller ID all assist in advanced search technology. However, a transcription process typically divorces this meta-data from the original recordings and reviewers lose this valuable asset to their search routines.

Regardless of the method used to review audio content, intelligent human listening is necessary to comprehend the body of content and develop the appropriate search and review techniques. And there is at least one method available that helps automate this process by making audio content searchable but still allowing review of the actual recordings.

One recent case demonstrates the benefits of using this automated approach. A transportation company involved in a significant litigation matter identified 350 hours of recorded telephone conversations consisting of 10,500 individual recordings from the company’s internal trading desk.

The company’s call-logger system was capable of exporting the files and metadata, including the date, time, duration, and identity of the company employee in each recording. This information was critical in reducing the amount of audio requiring review. Utilizing a combination of keyword and metadata searches, a single attorney was able to isolate 360 recordings (35 hours) that were responsive at a cost savings of 75%.

Conclusion

Intelligent audio search, one that combines metadata with thoughtful informed review, is the only reasonable and cost-effective audio discovery solution for litigation, regulatory and internal investigations. And when dealing with audio content, keep in mind the three differentiators – standardization, formality, and change – between the spoken and written words to get the best results.

Jeff Schlueter,
is Vice President,
Legal Market
for Nexidia, Inc.,
the leading provider
of phonetic
audio discovery.
Mr. Schlueter has an
extensive background
in audio technology.
Prior to Nexidia he
was Executive Director
of Voice Product
Management at
Cingular Wireless where
he was responsible
for development and
lifecycle management
of all voice related
products and services in
the Cingular portfolio.
He is also the founder
of The Messaging
Industry Association
(tmia.org), an industry
group dedicated to the
development and use of
electronic voice services.