Audio analytics: new opportunities in litigation and investigation
In today’s legal and regulatory environment, companies understand the importance of information governance, monitoring programs, retention requirements and collection processes as applied to electronic records such as email, instant messages, user files and financial data. But what about audio data? Until recently, audio data has not received much attention from regulators. But that is rapidly changing. Ask yourself: how much do you know about your company’s audio data? Where is it stored? In what format is it stored? Most importantly, if your company receives a regulatory demand, would your organization know what was in the computers’ audio files? Would your organization be prepared for an investigation?

With the right tools and strategy, technology-based audio analytics can be indispensable not only in addressing regulatory compliance, but also in promoting best practices and reducing potential settlement costs.

We will first review the rapidly expanding role of audio data in regulation and litigation. We will then argue that there are several key advantages of technology-based audio analytics that combine the insights of computational linguistics and machine learning with traditional industry expertise. Finally, we will offer strategies and practices on how to use leading edge technologies in managing — and reaping new benefits from — the complex landscape of audio data.

Growing prominence of audio data in regulation and litigation

Major government fraud investigations are increasingly turning their focus to audio data. In recent years, the U.S. Consumer Financial Protection Bureau (CFPB) and U.S. Commodities Futures Trading Commission (CFTC) have been especially active in analyzing phone recordings and in creating rules about the retention of this data type.

The CFPB, whose mandate is to make and enforce rules for consumer finance markets, has shown particular interest in audio analytics. In its winter 2015 Supervisory Highlights report, the bureau presents two cases involving speech data. 1 In the first case, the CFPB indicates that consumer reporting agencies (CRAs) are not meeting their dispute handling obligations under the Fair Credit Reporting Act (FCRA) by failing “to forward relevant documents submitted by consumers, including cancelled checks, invoices, and correspondence, to furnishers.” 2 In other words, the CRAs were insufficiently proactive in alerting debt collectors with debtors’ information. In working with the CRAs, the report points to “call center scripts and training regarding solicitation of relevant information from consumers with disputes.” 3 This investigation reveals that the CFPB has the authority to investigate recorded call center conversations, scripts and processes. 4

The second report in the 2015 Supervisory Highlights concerns supervision of creditors collecting due payments. 5 Of interest were “false and misleading representations in debt collection communications.” 6 The CFPB found that certain phone representatives were overly optimistic about the rehabilitation of debtors’ credit scores, misrepresented the waiving of collection fees and misled debtors into believing they must pay electronically. 7

In this matter, audio data was specifically reviewed and addressed by the CFPB, confirming that audio analytics is clearly within the CFPB’s regulatory toolkit. It can and will put audio data at the center of its investigations.

The second federal trade organization leveraging audio analytics is the CFTC. Its most significant recent fine, $115 million against Barclays in May 2015, was based on allegations that Barclays traders were manipulating ISDAfix, a reference for fixed interest rate swap rates that are collected daily by the International Swaps and Derivatives Association. Audio recordings of the traders with alleged misconduct were key to the case: the CFTC reported that “it found emails and audio recordings that were evidence of rigging.” 8 This matter offers an example of the liability that recorded speech data creates, yet also the value that proactively monitoring employees for compliance can generate. The CFTC ordered Barclays not only to cease and desist from additional violations, but also

---

2 Ibid, p. 5.
3 Ibid, p. 5.
5 “The Bureau’s recent examinations have identified a risk of a deceptive practice, violations of the FCRA and Regulation V, and violations of the Fair Debt Collection Practices Act (FDCPA),” – CFPB, Supervisory Highlights, Winter 2015, p. 6.
7 For more about this investigation into the Dept of Ed., see USC 1692e, “CFPB Examination Procedures, Debt Collection”. For articles about this debt collection supervision, see: Tim Bauer, “Next Step in CFPB Debt Collection Rulemaking,” InsideARM.com, 12 May 2015; “Consumer Lawsuits and CFPB Complaints Against Debt Collectors Increase in March,” ACA International; Chad Samet, “Debt Collectors Need CFPB’s Kick in the Pants,” American Banker, 28 April 2015.
to “take specified remedial steps, including measures to detect and deter trading intended to manipulate swap rates such as USD ISDAfix.”

In acknowledgement of the potential of speech data, in 2012 the CFTC mandated that all derivatives traders keep audio records. In the Barclays inquiry, the CFTC expanded the obligation to include not just retention, but also monitoring.

There have been other major fines handed out in decisions which rest, in part, on audio speech data. In April 2015, Deutsche Bank was fined $2.5 billion by the UK Financial Conduct Authority (FCA) for traders who were caught manipulating London Interbank Offered Rate. In this case, the bank’s faulty phone recording system failed to adequately allow monitoring of its traders. In another case not strictly related to speech text but to audio surveillance, AT&T was fined $25 million by the FCC for consumer data theft in international call centers.

As audio recording technology has grown, so too has its application in legal and regulatory investigations, and the fines levied by the regulatory entities in these cases are significant. This trend suggests that speech analytics will become increasingly important in monitoring, promoting and evaluating the legal behavior of corporations and employees, and will be essential to avoiding and reducing costly financial penalties.

Advantages of technology-based speech analytics over manual review

Given this fast-growing and increasingly sophisticated demand for audio data insights, choice of analytics methodology is paramount. Technology-based speech analytics solutions, leveraging insights from computational linguistics and machine learning along with strong regulatory expertise, can far outpace traditional, manual review of audio data by human listeners and offer unique new opportunities in regulatory and discovery applications. Indeed, technology-based speech analytics is already beginning to bear fruit in the consumer space, where large-scale audio analytics are used to increase call center efficiency, streamline market research and even proactively manage risk. For example, technology-based speech analytics is helping enforce call script

---

10 According to Bloomberg Business, “The rule will make enforcement investigations more efficient by preserving critical evidence that otherwise may be lost to memory lapses and inconsistent recollections,” CFTC Chairman Gary Gensler said in a statement. “The commission will have access to evidence of fraud and market manipulation, which is expected to increase the success of enforcement actions for the benefit customers, market participants and the markets” – Silla Brush, “Audio Records of Swaps and Futures Trades Required by CFTC,” Bloomberg Business, 17 December 2012.
adherence and also helping detect fraud in real time in call center interactions.13 In the enforcement and eDiscovery space, however, some of the advantages of technology-based speech analytics over traditional manual review are even more compelling, as discussed below.

The key appeal of technology-based speech analytics over traditional methods is cost and efficiency. Elimitinating or reducing, even by a small proportion, the volume of speech data that human reviewers listen to can result in enormous cost savings. This is doubly true in the eDiscovery and enforcement space, where legally trained managed document review teams command high hourly wages. In addition, technology-based analytics is fast. Whereas human review can take upwards of one-and-a-half times the duration of the original audio to review, speech analytics algorithms can deliver insights virtaully instanataneously after setup and tuning. Technology-based analytics is also scalable: given an existing speech analytics workflow, the marginal cost to process enormous volumes of additional data is small. Indeed, given the inexorable advance of big data analytics, automatic speech analytics may soon become an industry norm rather than a discretionary initiative.

The lynchpin in this interaction is a choice between the two broad classes of algorithms within speech analytics: phonetic search and automatic speech recognition (ASR; also known as continuous speech recognition).15 In phonetic search, an algorithm determines the phonetic closeness between a search term and a raw audio stream. On the basis of a predetermined confidence threshold, the system then outputs a hit report listing the time-stamped occurrences, if any, of each search term item in each call. In contrast, an ASR speech recognition system converts a raw audio stream to a hypothesized natural language transcript, and any searches for specific terms are subsequently run over the resulting text transcript. This latter step can make use of standard eDiscovery workflows, search optimization and existing text analytics infrastructure.

Processing speed and accuracy of performance are, on average, relatively similar between the two approaches. While phonetic search was historically faster and less computationally intensive than ASR, this gap has closed substantially in recent years.16

These advantages of technology-based speech analytics, however, require that the technology be harnessed in the right way. No “one-size-fits-all” solution exists. Instead, several variables should inform choice of strategy, including legal or analytical goals, the specific strengths of different technologies within speech analytics, and the nature of the data. And choice of strategy will, in turn, inform the amount and type of human manual review required to support the technology-based analysis. The interaction of these variables and practical recommendations can help you choose the strategy that best fits your objectives.

Other advantages of technology-based speech analytics are more subtle, yet no less important in the legal space. In contrast to human reviewers, a speech analytics algorithm is consistent and objective: given its detection of a certain phrase in a call, the algorithm will always make the same classification for that call, according to exact parameters that can be inspected and modified. Manual review, on the other hand, suffers from an often-ignored subjective component: even with clear guidelines, each reviewer makes classifications under a different and virtually unknowable array of subtle biases and influences. Indeed, a meta-analysis of inter-annotator agreement studies revealed that human reviewer F-scores—a measure of the frequency with which different annotators made consistent classifications—rarely exceed 70%.14

A final advantage of technology-based speech analytics over human review, especially in proactive and discovery-focused cases, is the ability to reveal unexpected, data-driven insights. Computational models can identify correlations that elude individual human reviewers. This is because models can be designed with far fewer inherent biases and expectations, and they can consequently explore much larger spaces of hypotheses for potential patterns than can humans. Furthermore, a single computational model potentially has access to many orders of magnitude more data than a single human reviewer, or even many teams of human reviewers, thus more readily exposing novel patterns in the data.

Technologies, strategies and caveats

These advantages of technology-based speech analytics, however, require that the technology be harnessed in the right way. No “one-size-fits-all” solution exists. Instead, several variables should inform choice of strategy, including legal or analytical goals, the specific strengths of different technologies within speech analytics, and the nature of the data. And choice of strategy will, in turn, inform the amount and type of human manual review required to support the technology-based analysis. The interaction of these variables and practical recommendations can help you choose the strategy that best fits your objectives.

The lynchpin in this interaction is a choice between the two broad classes of algorithms within speech analytics: phonetic search and automatic speech recognition (ASR; also known as continuous speech recognition).15 In phonetic search, an algorithm determines the phonetic closeness between a search term and a raw audio stream. On the basis of a predetermined confidence threshold, the system then outputs a hit report listing the time-stamped occurrences, if any, of each search term item in each call. In contrast, an ASR speech recognition system converts a raw audio stream to a hypothesized natural language transcript, and any searches for specific terms are subsequently run over the resulting text transcript. This latter step can make use of standard eDiscovery workflows, search optimization and existing text analytics infrastructure.

Processing speed and accuracy of performance are, on average, relatively similar between the two approaches. While phonetic search was historically faster and less computationally intensive than ASR, this gap has closed substantially in recent years.16


Typical accuracy for a phonetic search system is 90% for broadcast news and 65% for conversation, while ASR accuracy is about 90% for broadcast news and 60% for conversation. However, the types of errors are very different across the two systems, necessitating different strategies to optimize performance given each system’s strengths and weaknesses.

A key difference between phonetic search and ASR, with important consequences for choice of search strategy, is the amount of reliance on linguistic context in each technology. Phonetic search uses very little linguistic context and instead simply matches an audio signal to sequences of speech sounds that are probable in the language. ASR, in contrast, generates transcripts by matching audio signal to sequences of whole words that are probable in the language. Thus, ASR is more context-sensitive and more dependent on a specific statistical language model. This is a double-edged sword: ASR systems perform well given a language model suited to the genre, dialect and specific vocabulary of the data to be analyzed, but they are more likely than phonetic search systems to be confounded by previously untrained vocabulary items and language varieties. On the other hand, linguistic context can be leveraged by an ASR system to improve its guesses, as well as by human reviewers using ASR output to speed review — whereas phonetic search neither leverages nor makes visible this kind of context, leaving a reviewer who is interested in contexts the sole option of manual listening around time-stamped hits. Lack of context sensitivity also means phonetic search is well suited for fuzzy phonetic matching of, for example, names and linguistic variants, by adjusting a simple confidence threshold. On the other hand, lack of context sensitivity makes phonetic search relatively poorly suited for short search terms, since different words and phrases often sound similar or identical when stripped of context.

These differences in technical approach lead to a robust generalization of the types of errors common to each system, which can inform search strategy. ASR casts a relatively narrow net, in that matches are found only when the exact search term appears in the transcript. Thus, fewer matches are found using ASR than in phonetic search, but they may be more likely to be correct matches. Phonetic search casts a broad net: many matches are found, but they are less likely to be correct matches. ASR, therefore, is well suited for use cases where an exact term must be present, while phonetic search is optimal for cases where any of a variety of terms satisfies a criterion. The two systems can also be combined in an ensemble approach: calls can be classified as matching a given search term when both systems find the term in the call (resulting in more stringent classification) or when either system finds the term in a call (broader coverage).

---

To illustrate choice among these technologies based on the goals of the analysis, consider the following hypothetical. A large bank waives the monthly fee for its premium checking account in months when a direct deposit is made to the account. However, an internal review reveals that in some calls, the customer service representative did not mention the direct deposit requirement, and the bank wishes to identify all such calls in order to proactively correct the misleading interaction with the customer. This presents at least three possible call analytics strategies:

- **A fully automated classifier.** An algorithm classifies calls according to whether a speech recognizer detects the phrase “in months with a direct deposit.” The algorithm is iteratively improved and validated through human review of a small, representative sample of calls.

- **Partial manual review.** The classifier mentioned above is used, but all calls in which the phrase was not detected are reviewed and classified manually.

- **Prioritized manual review.** An algorithm ranks all calls according to the confidence with which keywords such as “direct” and “deposit” are detected. All calls are sent for manual review, starting with those where the system reports lowest confidence that the keywords were present.

Choice among these strategies depends on factors such as the goals of the analysis (determined, for example, by the precise regulatory requirements at play), data quality and system performance, as discussed above.

As the relative intricacy of the choice of search strategy has suggested, a few words of caution are warranted in the use of technology-based speech analytics. The approach is not a panacea for all regulatory and discovery matters and does not entirely eliminate the need for human review and domain-specific expertise. Data quality and technological limitations sometimes limit the use of automated analytics. For example, the audio quality may be too poor to be reliably parsed by speech recognition engines, or situations where training data for the systems’ language models does not match the target data to be analyzed, which can lead to unacceptably poor performance. In these situations, it may be prudent to take a conservative route and only eliminate from human review a (possibly very small) portion of calls for which confidence is high.

The caveat is that even in the best of conditions, human review of some calls is necessary to develop classification logics, tune speech analytics models and validate the models’ output, as the three search scenarios above illustrate. Human reviewers can identify specific shortcomings and common errors of early versions of speech analytics workflows, which can then be iteratively revised until performance of the automated system reaches an acceptable threshold. Human validation of an appropriate, statistically valid sample of calls is needed to verify that the analytics system is performing as expected.
Just as they do not eliminate the need for human review, technology-based speech analytics solutions do not replace the need for domain-specific expertise in enforcement, discovery and fraud detection. This is especially true given that most speech analytics technologies are not packaged as targeted applications for specific use cases, but rather as toolkits that can be leveraged in larger workflows.

Conclusions
With increasingly frequent use of audio-related data in enforcement actions by regulatory agencies, audio data is establishing itself as a key component of the investigative process, and strategies for analyzing this data are becoming indispensable. Although the landscape is complex, ever-maturing technologies and increasingly refined methodologies and search strategies are making audio analytics not only viable as part of a reactive strategy, but indeed desirable as part of a proactive toolkit to lessen the probability and optimize the outcome of litigation. In this way, the right combination of technology, strategy and human expertise will allow for the successful exploitation of the new world of speech analytics.

Contacts
Michael Faske
Partner, Head of Fraud Investigation & Dispute Services Switzerland
Tel: +41 58 286 3292
michael.faske@ch.ey.com

Paul Wang
Partner, Head of Forensic Technology & Discovery Services Switzerland/Regional Head Swiss Romandie of Fraud Investigation & Dispute Services
Tel: +41 58 286 5826
paul.wang@ch.ey.com

About EY
EY is a global leader in assurance, tax, transaction and advisory services. The insights and quality services we deliver help build trust and confidence in the capital markets and in economies the world over. We develop outstanding leaders who team to deliver on our promises to all of our stakeholders. In so doing, we play a critical role in building a better working world for our people, for our clients and for our communities.

EY refers to the global organization, and may refer to one or more, of the member organizations of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a UK company limited by guarantee, does not provide services to clients. For more information about our organization, please visit ey.com.

About EY’s Fraud Investigation & Dispute Services
Dealing with complex issues of fraud, regulatory compliance and business disputes can detract from efforts to succeed. Better management of fraud risk and compliance exposure is a critical business priority — no matter the industry sector. With our more than 4,200 fraud investigation and dispute professionals around the world, we assemble the right multidisciplinary and culturally aligned team to work with you and your legal advisors. And we work to give you the benefit of our broad sector experience, our deep subject matter knowledge and the latest insights from our work worldwide.

© 2016 Ernst & Young LLP.
All Rights Reserved.

1508-1600162 NE
SCORE #: WW0417
ED None

This material has been prepared for general informational purposes only and is not intended to be relied upon as accounting, tax, or other professional advice. Please refer to your advisors for specific advice.

ey.com