NICE Speech Analytics
Why Combining Phonetics and Transcription Works Best
An Overview of the Industry’s First Hybrid Speech Analytics Technology
# TABLE OF CONTENTS

What is Speech Analytics? 1

Common Methodologies 1

A Hybrid Approach 1
- Quick Time to Insight 1
- 100% Call Coverage 1
- Accurate Call Categorization and Search 2
- Root-Cause Analysis 2

Real-Time Speech Analytics 3

Additional Speech Analytics Building Blocks 3
- Emotion Detection 3
- Speaker Separation 3
- Talk Analysis 4
- Call Part Analysis 4
- Additional Data Sources 4

Why NICE Speech Analytics? 4
What is Speech Analytics?

Speech analytics refers to automated methods of analyzing speech to gain greater insight into customer interactions, and business and individual agent performance. Speech analytics applications are commonly deployed in contact centers, where a large number of calls take place every day between customers and customer service representatives. Speech analytics goes beyond identifying spoken words. It applies linguistic and semantic analysis to verbal conversations in order to understand the topics discussed and their context, and the sentiment of the speakers during the interaction. Speech analytics is usually part of a broader interaction analytics solution that extracts the voice of the customer from multiple channels such as phone, email, chat and social media.

Common Methodologies

Every speech analytics process starts by identifying phonemes, which are the most basic units of speech. Phonemes are the building blocks of words, and are different for each language. There are different methodologies for translating identified phonemes into contextual insight. Two main methodologies dominate the market:

- **Phonetic indexing and search**: This approach provides the fastest time to insight. Since most languages only have a few dozen unique phonemes, identifying and indexing them enables quick and accurate search for words and phrases, and easy categorization of calls to various topics, even within a large number of calls. Using this methodology, users can search for any word or phrase regardless of whether it appears in a dictionary (most useful for product names), and quickly view trends in call categories.

- **Transcription**: Also known as Large-Vocabulary Continuous Speech Recognition (LVCSR) or speech-to-text, this approach hinges on the automated recognition of known words. It is much slower than phonetic indexing since a typical language has tens of thousands of words in its dictionary, and each spoken word within the analyzed audio needs to be identified among this very large group of candidates. However, it enables data mining and natural language processing to automatically surface the root causes of unknown issues.

A Hybrid Approach

Rather than choosing between phonetic indexing and transcription, NICE Speech Analytics combines both speech methodologies into a software solution without equal. It allows contact centers to quickly categorize and trend 100% of their calls, and to understand the issues at the heart of each and every one of them. NICE is the first to market with this kind of powerful hybrid approach.

How does this novel NICE methodology work? How does hybrid speech technology benefit call centers? Here, we explain.

Quick Time to Insight

Categorizing calls is fundamental to speech analytics, and is the first step to understanding the broad context of customer interactions. As calls take place, NICE Speech Analytics automatically categorizes them as billing calls, claims calls, cancellation calls, repeat calls, dissatisfied customer calls and so on. Trending call categories across time can identify problems, such as a spike in calls about service (perhaps there was a sudden outage?) or an acceleration in customer dissatisfaction (maybe the new format for billing statements is confusing?), so the quicker time to insight, the better. Since phonetic indexing is significantly faster than full transcription, it is the most effective technology to power call categorization. Speech analytics systems that do not support phonetic indexing may take four times longer to get to the same insights compared to systems that leverage phonetic indexing.

100% Call Coverage

Another benefit of phonetic indexing is its capacity to analyze entire call volumes. Since transcription is an intense draw on CPU usage, a single transcription server can analyze a significantly lower number of calls, more slowly, compared to a phonetic server. Thus, solutions that rely solely on transcription are unable to process large call volumes. In fact, in many cases, they can only analyze a random sample of calls. This limitation may have a serious impact on the resulting business insight. For example, a solution that randomly analyzes 20% of calls will only identify 20% of customers at churn risk; the other 80% of customers who may defect to another provider will remain “off the radar.” NICE Speech Analytics covers 100% of calls.
Accurate Call Categorization and Search

Transcribing a spoken interaction is inherently inaccurate. It is subject to several factors such as speaker accent, speech speed and quality of the call connection. Plus, since transcription relies on the dictionary of the interaction language, every word spoken must be identified out of tens of thousands of possible candidates. Since many words sound similar if not identical, it is statistically likely that some words will be transcribed incorrectly. For example, the word “error” may be incorrectly transcribed as “err or” or even “hair.” In fact, research shows that the Word Error Rate (WER) in English-language call center interactions exceeds 50%\(^1\). This means that, on average, every other word is transcribed incorrectly. The WER is even higher for other languages, for which the language models are not yet as mature as they are for English.

Categorizing calls based on inaccurate transcriptions inevitably leads to inaccurate results. Many calls fall through the categorization cracks as their context signals (i.e., words) are missing or wrong. In addition, any ad hoc search for long phrases will often fail to identify the calls in which the given combination of words was said, since 50% or more of the words transcribed are false.

Phonetic indexing is a much more accurate method of categorization. It has the advantage of using a small pool of only a few dozen phrases per category, compared to tens of thousands of dictionary words that a transcription process has to choose from. In addition, ad hoc search based on a phonetic index is more effective since it can find any word, even words that are not in the dictionary, such as product names. Transcription, on the other hand, is limited to a finite list of dictionary words.

Root-Cause Analysis

Once an issue is identified, such as a spike in repeat calls, root-cause analysis is needed to get to the bottom of the problem. For example, analyzing what customers say during repeat calls will uncover important clues as to why their original issues went unresolved.

Solutions based on phonetic indexing alone fall short in this area because their search function looks for predefined words and phrases. But in many cases, users may not know what to look for. Without advanced data mining capabilities covering the entire conversation, manually listening to random calls may be the only way to hunt for clues.

NICE Speech Analytics features root-cause analysis based on transcription—and this is where this methodology shines. As calls are categorized, transcription draws on the following analytical tools to generate a topics list (See Figure 1) for root-causes of problems:

Linguistic Analysis

Linguistic analysis leverages natural language processing technologies to surface logical root-cause topics. For example, the phrases “too expensive” and “two expensive” sound alike, so they may be transcribed either way. However, since “too expensive” is a more valid and likely syntactic form than “two expensive,” the latter will not be considered as a root-cause.

Statistical Analysis

Statistical analysis is used to calculate how salient a phrase is within a call category. For example, the phrase “service provider” might be incorrectly transcribed as “service provide her.” But since “provide her” is unlikely to be unique to a certain call category, it won’t be featured in the root-cause topic list, while “provider” might be.

Figure 1: Root-Cause Analysis Automatically Surfaces Customer Issues

---

\(^1\) https://www.research.ibm.com/haifa/Workshops/speech2008/program.shtml
Context Visualization

Once a root-cause phrase is identified (e.g., “problem with my payment”), the user can perform context analysis to identify related words and phrases that will help shed more light on the issue. NICE Speech Analytics displays this analysis visually, as shown in Figure 2. The size of a bubble and the thickness of the line connecting the two phrases represent how closely the phrases are related, and how significant the phrases are within the call category in question.

Real-Time Speech Analytics

Is there really a difference between analyzing a call as it happens, and analyzing a recorded call an hour later? Absolutely. In fact, some insights are most valuable while the call is taking place. Interest in an additional product or service, or signals of potential churn or even fraud are insights that must be acted on in real time. By the time the call is over, a sales opportunity may be missed, a customer lost or a fraudulent transaction processed.

NICE Real-Time Speech Analytics is a new addition to the NICE Speech Analytics platform that identifies actionable insights in real time. It analyzes calls as they unfold, rather than analyzing call recordings later. As insights are revealed, real-time on-screen guidance assists agents with next-best-action recommendations, helping them to deftly handle the customer situation at hand. Real-time alerts also can be sent to supervisors when appropriate.

Additional Speech Analytics Building Blocks

Emotion Detection

As the saying goes, “it’s not what you say, but how you say it.” Tone can make all the difference. For example, a customer may make a positive comment, “You have great service” or an ironic, angry statement “Oh, you have GREAT service, alright!” The words are nearly identical, but certainly not their meaning. It takes sophisticated speech analytics to know the difference. Emotion detection is crucial in order to truly understand the voice of the customer.

The emotional state of a speaker can be identified by features such as variants of pitch, energy, prosody (patterns of stress and intonation) and spectral features. During emotion-rich segments of an interaction, the statistics concerning these features will differ markedly from those found during periods of neutral speech. NICE Speech Analytics identifies all calls in which speakers exhibit a high level of emotion. When playing back a call, emotional events will be visually marked on the playback screen (see Figure 3), allowing a supervisor or quality manager to quickly skip to areas of interest within the call, saving time and effort.

Speaker Separation

In order to fully understand the context of verbal statements and to use the insights generated from speech analytics most powerfully, it is important to identify who is speaking at key points in the conversation. Consider a scenario where the word “cancel” is said in relation to a customer’s service. If a customer says, “please cancel my account,” that’s vastly different from an agent saying, “Thanks for signing up. You have 30 days to cancel.”

Speaker separation can be implemented at the telephony-infrastructure level, where the agent and customer channels are recorded separately. However, this configuration is not always supported by the telephony environment or may not be practical. NICE Speech Analytics offers software-based speaker separation. It uses sophisticated acoustic algorithms to separate two speakers on a single audio channel into two virtual channels, allowing their speech to be analyzed discretely.
Talk Analysis

Tracking the give-and-take of interactions can yield useful insights into agents’ abilities. A situation where an agent and customer talk over one another typically indicates poor call-handling skills. Similarly, an agent’s excessive silence may indicate a knowledge gap, especially when it occurs during calls of the same type. Silence also can indicate that an agent is “muting” the phone, which can be hard to detect otherwise. The talk analysis function in NICE Speech Analytics provides insight into consecutive talk time of customers and agents, silence, talk-over time, and the number of bursts where one speaker is interrupting the other—all of which can indicate opportunities for additional training or coaching.

Call Part Analysis

Call part analysis is an innovative technology that leverages natural language processing to separate transcribed calls into unique parts. For example, a typical interaction may flow from customer identification to problem description and resolution to up-sell and wrap-up. By identifying these discrete parts within calls and calculating the handle time for each, the system can identify, for example, which agents are not well-trained in customer verification procedures, or which agents are not spending enough time on up-sell attempts. NICE Speech Analytics displays call part analysis graphically (see Figure 4) so supervisors can easily understand what's driving handle time and whether agents are spending call time efficiently.

Additional Data Sources

The insights derived from speech analytics are even more powerful when complemented by other data sources that highlight additional dimensions of interactions. NICE Speech Analytics integrates with computer telephony integration (CTI) to expose call holds or transfers during interactions. Desktop analytics provides insight into how agents are using their computers—accessing applications, making keystrokes—while on calls with customers. Integration with desktop analytics can uncover knowledge gaps (lack of familiarity with certain applications) or inefficient processes that extend call handle time unnecessarily. It also makes it possible to attach to interactions data from CRM systems and other applications on the agent's desktop such as customer lifetime value, demographics, transaction history and so on, which may be useful for further analysis.

Why NICE Speech Analytics?

NICE Speech Analytics is like no other solution available today. Its hybrid approach combines phonetic indexing with transcription so there’s no need to choose between quick time to insight and root-cause analysis, or to base business decisions on just a small sample of random calls. This sophisticated real-time speech analytics solution, including talk and call-part analysis, speaker separation, emotion detection and more, supports more than 20 languages and dialects, including English, French, German, Italian, Japanese, Spanish, Portuguese, Russian, Turkish and Hebrew, to name a few. And as part of NICE Interaction Analytics, it shares the same methodology of categorization and root cause analysis with NICE Text Analytics, extending our best-of-breed analytics technology to text-based channels such as email and chat. Together, NICE Interaction Analytics is a powerful platform to uncover insights from interactions and enable real-time business impact.

“NICE is in a unique position in the speech analytics market, as it offers both LVCSR and phonetic indexing, realtime and post-call speech analytics. It offers these alongside process analytics, customer feedback tools, and performance management to create a full WOTs suite. Its customers can view both customer needs and agent performance across different channels from within its application.”

Aphrodite Brinsmead, “Realtime Speech Analytics in the Contact Center”, OVUM, July 2011
NICE Systems (NASDAQ: NICE), is the worldwide leader of intent-based solutions that capture and analyze interactions and transactions, realize intent, and extract and leverage insights to deliver impact in real time.

Driven by cross-channel and multi-sensor analytics, NICE solutions enable organizations to improve business performance, increase operational efficiency, prevent financial crime, ensure compliance, and enhance safety and security.

NICE serves over 25,000 organizations in the enterprise and security sectors, representing a variety of sizes and industries in more than 150 countries, and including over 80 of the Fortune 100 companies. www.nice.com.

Disclaimer

Information in this document is subject to change without notice and does not represent a commitment on the part of NICE Systems Ltd. The systems described in this document are furnished under a license agreement or nondisclosure agreement.

All information included in this document, such as text, graphics, photos, logos and images, is the exclusive property of NICE Systems Ltd. and protected by United States and international Copyright laws. Permission is granted to view and photocopy (or print) materials from this document for personal, non-commercial use only. Any other copying, distribution, retransmission or modification of the information in this document, whether in electronic or hard copy form, without the express prior written permission of NICE Systems Ltd., is strictly prohibited. In the event of any permitted copying, redistribution or publication of copyrighted material, no changes in or deletion of author attribution, trademark legend or copyright notice shall be made.

360o View, Alpha, ACTIMIZE, Actimize logo, Customer Feedback, Dispatcher Assessment, Encorder, eNiceLink, Executive Connect, Executive Insight, FAST, FAST alpha Blue, FAST alpha Silver, FAST Video Security, Freedom, Freedom Connect, IEX, Interaction Capture Unit, Insight from Interactions, Investigator, Last Message Replay, Mirra, My Universe, NICE, NICE logo, NICE Analyzer, NiceCall, NiceCall Focus, NiceCLS, NICE Inform, NICE Learning, NiceLog, NICE Perform, NiceScreen, NICE SmartCenter, NICE Storage Center, NiceTrack, NiceUniverse, NiceUniverse Compact, NiceVision, NiceVision Alto, NiceVision Analytics, NiceVision ControlCenter, NiceVision Digital, NiceVision Harmony, NiceVision Mobile, NiceVision Net, NiceVision NVSAT, NiceVision Pro, Performix, Playback Organizer, Renaissance, Scenario Replay, ScreenSense, Tienna, TotalNet, TotalView, Universe, Wordnet are trademarks and/or registered trademarks of NICE Systems Ltd. All other trademarks are the property of their respective owners.