Whitepaper
Speech Analytics Tools

02/2017
1 Introduction
Hearing the voice of the customer presents a challenge to even the most sophisticated contact center. Many different measurements are used to determine and evaluate the service quality of customer interactions such as disconnection rates, holding times or reaction times. But these measurements tell you about events within customer interactions instead of the reason why they occurred.

Here speech analytics comes into play. With speech analytics, you can automatically identify and extract relevant information from unstructured data for an analysis impossible to conduct manually in a cost-effective way. By using speech analytics, you can optimize business processes or agent coaching and boost customer satisfaction continuously.

Just five years ago, speech analytics was thought of like "rocket science," while today, it is a well-tested and accepted tool to significantly improve the quality of contact center interactions. This paper describes the pros and cons of the major approaches to speech analytics and its benefits for contact centers as well as providing tips for successful implementation.

2 Speech Analytics Technologies
Speech technologies are used to extract information from spoken data, but professional systems must be speaker independent because the text processing software you use at home to write a letter can be especially trained for your voice. In addition, at contact centers, the various callers speak without any long pauses, making it difficult to separate the words.

To meet these challenges, speech technologies utilize phonemes, the smallest units of voice. Every language contains 10-to-80 typical phonemes. Phoneme analysis compares the given voice frequencies with the known frequencies of phonemes in the language. By using the known probabilities of typical phoneme combinations, the accuracy of the transcription can be increased.

3 Speech Analytics Methods
The three basic speech recognition approaches, keyword and phrase spotting, phonetics and LVCSR (large-vocabulary continuous speech recognition), differ in the type and depth of analysis, the effort needed for preparation of the system, and the startup costs.

3.1 Common but Effective: Keyword and Phrase Spotting / Phonetic Indexing
Keyword and phrase spotting provide as well as phonetic indexing are common and cost-effective possibilities to filter or categorize calls based on specific words or phrases. Both approaches are based on the analysis of phonemes, that is to say speech is divided into its smallest units – the phonemes. Thereby, a comparison of the phonemes of a search term and the phonemes of the audio file is possible.

The speech is stored as an audio file, and you search based on the way a word sounds, not how it is spelled. This method facilitates searching for obscure words or those with unknown spelling such as names or places. Simultaneously, the system might produce a false positive depending on the context as well as the existence of homophones and homonyms. (Homophones describe words pronounced in the same way with different spellings and meanings, e.g., “for” and “four,” while homonyms share the same spelling and pronunciation with different meanings, e.g., “left” as the past tense of leave and “left” as the opposite of right.).

Both approaches can be easily implemented with enormous benefits. Selective searching for critical calls can identify issues for agent improvement and provide ideas for workforce optimization.

While keyword and phrase spotting is based on the processing of pre-defined words and phrases, phonetic indexing captures every single word of a conversation within an index. Both methods have their pros and cons.

Keyword Spotting
Voice recognition is processed only for specific criteria, namely the words or phrases you set up in a keyword list. Since the system does not interact with speech data as a whole, the approach is limited to the recognition of previously defined words and phrases. New keywords require a new analysis of the audio material. So, the keyword and phrase spotting approach is preferred for identifying known issues which do not change very often.

Phonetic Indexing
BWith a phonetic approach, identified phonemes of a call are stored and indexed in a database, with the phonemes of the search term then matched against the phonetic index. Repeat searches are expedited once a phoneme translation has been completed. The user does not need to repeat the analysis of speech data for every new search value because this approach is not limited to pre-defined words. Phonetic indexing is indicated for filtering calls on the basis of flexible word lists.

Compared to keyword and phrase spotting, processing of words is slower and more memory capacity is needed.
3.2 Speech-to-Text Transcription: Large-Vocabulary Continuous Speech Recognition (LVCSR)

LVCSR converts the call into text by setting up a language model using dictionaries with hundreds of thousands of words. These dictionaries are used to interpret the audio, thus slowing down the processing speed compared to phonetic indexing. But after the complex speech processing is completed, words can be found quickly with a text-based search.

Of course, this approach surpasses phonetic indexing and keyword and phrase spotting by providing a textual version of the call. Any thinkable word can be searched, and words can be seen within their context. For further analysis, data can be transferred to other systems (e.g., a data warehouse system) where it can be easily used to explore additional trends or events. In general, an LVCSR system requires extensive preparation and training, but its ability to use data for any type of analysis makes it the most powerful solution.

4 Further Analysis: Emotion Detection

Emotion detection broadens the possibilities of traditional approaches to speech analytics because it is not based solely on the words used. Everyone expresses themselves differently; thus, words and feelings do not always match up. Therefore, emotion detection concentrates on changes in one’s voice.

Acoustic and prosodic characteristics are analyzed to find out whether a person is angry, neutral or happy. If a person suddenly starts screaming at an agent, the system recognizes the change in volume. Different levels of an emotion, such as “very angry” or “lightly angry,” can be detected as well as various emotions such as anger and happiness. However, the number of emotions analyzed increases the probability of errors. But with constant adjustments, the system’s recall and precision might rise to 90 percent.

5 Benefits of Speech Analytics

As everyone knows, the amount of daily business communications is huge. Think about all the incoming and outgoing letters, emails and phone calls an organization must handle. And don’t forget internal communications. The information contained in all these communications is often critical for workforce optimization in other company departments beyond the contact center.

But who can separate out and analyze the relevant information? For humans, this would involve a time-consuming, never-ending and costly process, but speech technologies can automate the filtering and analysis. Speech analytics enables the monitoring of conversations, making it both easier and more effective. A huge quantity of data can be evaluated, with relevant calls extracted for review and flagged for analysis. The flexibility of speech analytics broadens its use from contact centers to the entire enterprise.

Please see below some of the key benefits:

• **Advancing training and coaching:** By filtering critical calls with speech analytics, the supervisor of a contact center might recognize agents’ problems with up-selling, for example, and then assign a seminar to improve their skills. The agents will benefit both in new skills and personal fulfillment, reducing turnover and improving customer interactions.

• **Improving business processes:** Speech technologies can evaluate critical processes. For example, if agents must deal with too many applications during a customer interaction, the customer might become impatient. By identifying these calls, a less time-consuming and expensive solution can be found, e.g., a CRM system could be integrated with only one central access point for all applications. Improving business processes can save money and man-hours that can then be directed to other essential areas.

• **Adhering to compliance requirements:** With speech analytics, 100% of calls can be verified as compliant. For example, when recording is used for quality monitoring, the adoption of a double opt-in method will provide valid documentation of customer agreement. Speech recognition can identify every call where customers agreed or denied permission to record. This way, you can avoid fines and unnecessary expenses.

• **Improving contact center indices:** If an important contact center benchmark indicates a problem, you can use speech technology for a root-cause analysis to find the responsible calls. After detecting the cause, you can take further steps. For example, a high average handle time may be caused by agents who must login for every application. As a solution, you could introduce a single sign-on method.

• **Hearing the voice of the customer:** Get a deeper insight into customer experiences by analyzing why they call and their problems, and how to best meet their needs.

• **Controlling service costs:** Do your agents spend too much time with basic customer questions and problems? Filter out these calls and try to resolve them without agent interactions, perhaps with the help of an IVR, thus giving agents more time for customers who really need them.

• **Getting a deeper insight into markets and business intelligence:** Filtering calls for customer satisfaction or dissatisfaction with your product or your competitors’ can help you spot market trends.
• **Avoidance of migration of customers**: Find angry customers with emotion detection and solve their issues before they get out of hand.

6 Introduce Speech Analytics into your Contact Center

Before introducing speech analytics in your company, you should **formulate your business needs**. Choose an area of improvement and benchmark it. Speech analytics can be used for many different purposes, so you must define how it will be used for your company.

You also should consider **technical constraints and cultural preferences** such as union requirements.

Here are some items you should consider when implementing a speech analytics solution:

Here are some items you should consider when implementing a speech analytics solution:

- Ask your vendor whether you will require **additional hardware** for the speech analytics engine.
- Look for a **flexible and scalable system** to meet your individual needs and expand if your requirements grow.
- The system should be **easy to configure and to adjust**.
- **Experiment** with your speech analytics solution before using it on a daily basis.

- Decide whether you want to **analyze past calls** to improve operations or whether you want to conduct **real-time speech analytics** for current interactions. A combination of both will provide a powerful tool to learn from past failures and intervene in urgent situations, perhaps if a customer threatens to switch to a competitor.
- Many vendors will conduct a **proof of concept** prior to final implementation so you know what to expect.
- During the **roll-out process**, you can clarify whether the solution really works, discuss technical details, such as the language model, and begin a process of testing and refining before training your staff and using speech analytics on a day-to-day basis.
- **Supporting various languages and dialects** requires different models and dictionaries. For similar accents, the same language model can be used. Your speech analytics solution should be able to handle different languages in the same system.
- Speech analytics accuracy will never reach 100 percent, but you will get enough information for an insightful analysis. **Recall and precision** work against each other so you will never reach total accuracy for both. (Precision describes the part of relevant information within your results; recall describes the completeness of results in relation to all relevant data in your database.) Determine your preferred balance for the two.

7 Conclusion

When adopting a speech analytics solution, you must consider the purpose and select the appropriate speech analytics approach carefully. Careful determination of your speech analytics goals increases the likelihood of an enduring and long-lasting use of this capability.

You should also assess the time needed to properly create keyword lists, dictionaries or language models, depending on the approach you choose. Speech analytics will only be as good as the infrastructure supporting it. However, after appropriate preparation, implementation and a little patience, you will be amazed by how much you can benefit from the new information you can uncover.