Speech Analytics Data Reliability: Accuracy and Completeness

The Prerequisite to Optimizing Contact Center Performance and the Customer Experience
Summary of Contents (Click on any heading below to jump directly to that section):

EXECUTIVE SUMMARY
WHAT IS SPEECH ANALYTICS?
THE UTOPY APPROACH TO SPEECH ANALYTICS
• SEARCH
• DISCOVERY
• CATEGORIZATION

DEFINITION OF DATA RELIABILITY: ACCURACY AND COMPLETENESS
• ACCURACY
• COMPLETENESS (HIGH DETECTION RATE)

SPEECH ANALYTICS APPROACHES
• TRANSCRIPTION METHODS: SPEECH-TO-TEXT AND SPEECH-TO-PHONEME
• UTOPY’S PHRASE-BASED METHODOLOGY: THE MEANING OF SPEECH

ESSENTIAL FACTS
• TRANSCRIPTION METHODS PRODUCE INCOMPLETE DATA
• RESULTS FROM ACADEMIC STUDIES
• RESULTS FROM CUSTOMER TESTS
• UTOPY’S PHRASE-DRIVEN APPROACH: FIVE TO NINE TIMES MORE RELIABLE

THE IMPORTANCE OF COMPLETENESS (HIGH DETECTION RATES)
• AGENT PERFORMANCE OPTIMIZATION: NEEDS IDENTIFICATION DIFFICULT WITH LOW DETECTION RATES
• MEASURING THE IMPACT OF TRAINING, COACHING OR POLICY CHANGES: TIME IS MONEY
• DECISION MAKING AND USER ADOPTION: DECISION-MAKERS CAN’T RELY ON ANECDOTAL DATA
• COMPLIANCE: THE IMPACT OF MISSING “NON-COMPLIANT” EVENTS
• CUSTOMER SATISFACTION (CSAT) AND LOYALTY: THE ABILITY TO MAXIMIZE CSAT INCREASES WITH THE AMOUNT OF CUSTOMER SENTIMENT DETECTED
• THE IMPORTANCE OF CONTEXT: PHRASES PROVIDE CONTEXT TO THE SPOKEN WORD

CONCLUSION
Executive Summary

UTOPY's unique approach to Speech Analytics delivers business value in three different ways: search, discovery and categorization. Categorization delivers the highest business value of these three usages because it is the most capable of automatically monitoring and optimizing contact center performance. UTOPY's patented phrase-driven direct speech analysis method delivers far better categorization than any of its competitors due to the much higher level of data reliability which it delivers.

Data reliability refers to the accuracy and completeness of data. Most Speech Analytics vendors will quote the accuracy of their product, but neglect to mention the completeness of the data when the product is tuned to the claimed accuracy level. In the realm of Speech Recognition and Analytics, usually as accuracy goes up, completeness goes down. However, the patented phrase-based methodology employed by UTOPY for speech categorization produces data that is much more complete compared to “transcription” methods at high levels of accuracy.

The impact of this difference is multiplied if this data is to be used for the purposes of critical business endeavors such as Contact Center Performance Optimization, regulatory compliance, or minimizing customer churn. Contact Center Performance Optimization requires reliable identification of the specific topics that each agent finds difficult to handle, compliance requires reliable identification of rare occurrences, and minimizing churn requires reliable identification of customer sentiment. Only UTOPY’s phrase-driven approach, which is five to nine times more reliable than the transcription approaches utilized by every other Speech Analytics vendor, is sufficiently reliable to drive applications which optimize contact center performance and the customer experience.

What is Speech Analytics?

Speech Analytics software takes free-flowing conversations between companies and their customers, analyzes those conversations and structures them into information that can then be leveraged for business value.

The UTOPY Approach to Speech Analytics

UTOPY's unique approach to Speech Analytics delivers business value in three different ways: search, discovery and categorization.

Search

Search is the process of searching for words or phrases that have been spoken within calls, generally on an ad-hoc basis. Search is useful within many analysis workflows. Searches can be used for ad-hoc exploration of conversation content or to test hypotheses.

UTOPY delivers Speech-to-Text transcription using a Large Vocabulary Speech Recognition (LVCSR) engine. This Speech-to-Text transcription powers UTOPY’s text driven search engine, which enables ad-hoc word searches and exploration. Search results are sorted by relevancy and speech recognition accuracy.
**Discovery**
Discovery uncovers current trends or events occurring within conversations—trends or events which an organization may not have previously been aware of. Discovery can be useful for uncovering important events or issues before they become major problems for the organization.

UTOPY offers various discovery analytics which are powered by the UTOPY Speech-to-Text engine. These discovery analytics uncover informative current trends or occurrences within calls and proactively notify the organization of these current issues or topics.

**Categorization**
Categorization is the process of categorizing calls based upon the reason for the call, the topics that were discussed during the call, and/or any events which occurred during the call. Of the three Speech Analytics use cases which UTOPY enables, categorization delivers the highest business value because it discovers the call reasons, agents, products or processes which are driving performance on contact center Key Performance Indicators (KPIs). UTOPY categorizes calls based upon the phrases spoken within the calls. UTOPY's patented phrase-driven direct speech analysis method delivers far better and more valuable categorization than any of its competitors due to the much higher level of data reliability which it delivers.

**Definition of Data Reliability: Accuracy and Completeness**
According to the US Government Accountability Office\(^1\), “data reliability refers to the accuracy and completeness of computer-processed data, given the uses they are intended for.” In the realm of Speech Recognition and Analytics, “completeness” is measured by the “detection rate”, and usually as accuracy goes up, the detection rate goes down.

**Accuracy**
According to the American Heritage® Dictionary of the English Language, accuracy is defined as: “The ability of a measurement to match the actual value of the quantity being measured.” In the realm of Speech Recognition and Analytics, accuracy refers to the portion of results that were correctly recognized within a given result set.

The concept of accuracy can be illustrated in laymen’s terms by the following example. An executive in San Francisco needs to attend a meeting in New York. The executive calls the company travel agent, and asks the travel agent to e-mail him a list of all the flights from San Francisco to New York. The travel agent e-mails the executive a list of ten flights, but one of the flights is from Oakland to Newark, while the rest of the flights are from San Francisco to New York as the executive requested. Nine out of ten is 90% accuracy.

**Completeness (High Detection Rate)**
Completeness is defined as: “The state of being complete and entire; having everything that is needed.” In the realm of Speech Recognition and Analytics, completeness is measured by the detection rate, and

---

Page 4

www.utopy.com
detection rate refers to the portion of occurrences of a given event or word found by the system compared to the actual number of occurrences. Throughout the remainder of this paper, we use completeness and high detection rate interchangeably, as they have the same meaning.

To continue with our previous travel agency example, imagine that after the executive receives the list of flights from the travel agent, the executive sits down at his computer, goes to an online travel website, and searches for flights from San Francisco to New York. His search returns 18 flights from San Francisco to New York. The list from the travel agent had only 9 of those flights. The travel agent had only found nine out of the 18 flights from San Francisco to New York—a 50% detection rate.

Needless to say, data reliability is critical to sound decision making (see The Importance of Completeness for details). However, most Speech Analytics vendors neglect to mention their detection rates when discussing the reliability of their data, instead quoting their accuracy rates, even though their detection rates go down significantly as their accuracy goes up. Throughout the rest of this paper, we provide the results of various tests done by independent third parties which prove that transcription methods of Speech Analytics produce incomplete data.

Speech Analytics Approaches

Transcription Methods: Speech-to-Text and Speech-to-Phoneme
Most Speech Analytics products use Speech-to-Text or Speech-to-Phoneme (also known as Phonetic) conversion methods. In these approaches, a speech recognition engine transcribes the audio into collections of words or phonemes (the smallest discreet unit of human speech), and then a text mining or search engine attempts to spot keywords—or combinations of keywords—in the converted text or phonemes.

These approaches enable rapid ad-hoc searches and exploration of unstructured audio data, which is why UTOPY utilizes a Speech-to-Text conversion engine to power its speech search and discovery capabilities.

However, data is permanently lost during the conversion process due to the inherent limitations of current speech recognition and conversion engines, and the context in which the words were used (and therefore the business impact or outcome) is very difficult to capture using such a keyword-spotting approach.

UTOPY’s Phrase-Based Methodology: The Meaning of Speech
In contrast, the patented, phrase-driven UTOPY approach to speech categorization and analysis combines speech and business concept recognition into a single-step.

By directly recognizing and analyzing entire phrases, UTOPY’s patented method is automatically tailored to specific business needs and industry specific terminology, making it an order of magnitude more reliable at recognizing speech that is relevant in a business context than other solutions in the market today. Transcription approaches, which only recognize individual words or phonemes, must customize their entire language models in order to address such particular business needs.
Furthermore, entire phrases are much better than smaller recognition units at compensating for the mispronunciations, utterances and other speech disfluencies\(^2\) common in everyday speech.

The result is a holistic approach to precisely and completely understanding and analyzing the entirety of the interactions taking place between companies and their customers, and an exponential increase in data reliability and business value.

**Essential Facts**

**Transcription Methods Produce Incomplete Data**
Speech-to-Text and Speech-to-Phoneme products can be tuned to be relatively accurate, but their detection rates suffer as a result. Only UTOPY’s patented phrase-driven direct speech recognition and analysis technology is able to maintain a high detection rate at high accuracy levels, as Table 1 below illustrates.

<table>
<thead>
<tr>
<th>Speech Recognition Engine</th>
<th>Phrase-Driven</th>
<th>Speech-to-Text</th>
<th>Phonetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit of Speech Recognition</td>
<td>Phrases</td>
<td>Words</td>
<td>Phonemes</td>
</tr>
<tr>
<td>Typical Deployed Accuracy</td>
<td>85-95%</td>
<td>80-85%</td>
<td>80-85%</td>
</tr>
</tbody>
</table>

**Tuned to the same accuracy, what is the detection rate (completeness) out of 10 events?**

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>80% accuracy</th>
<th>80% accuracy</th>
<th>80% accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection (completeness)</td>
<td>70% detection</td>
<td>10% detection</td>
<td>5% detection</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10 events</th>
<th>9 results</th>
<th>1-2 results</th>
<th>0 results</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 events</td>
<td>7 true events found</td>
<td>1 true event found</td>
<td>0 true events found</td>
</tr>
</tbody>
</table>

Table 1: A comparison of the various Speech Analytics approaches

Table 1 assumes that 10 events have occurred within a set of calls. When tuned to 80% accuracy, the detection rate of UTOPY’s phrase-driven direct speech recognition and analysis method is 70%, the detection rate of UTOPY’s Speech-to-Text conversion method is just 10%; and the Phonetic method has only a 5% detection rate. Therefore, of those 10 events that have occurred, UTOPY’s phrase-driven method will accurately detect 7, while its Speech-to-Text method might only accurately detect 1, and the Phonetic method may not accurately detect any events!

---

Results from Academic Studies
Various academic studies have proven that transcription methods of Speech Recognition and Analytics have low data reliability. For example, according to a recent, well established study\(^3\), when used with real-world call center data, the Speech-to-Text method is only around 40-50% correct at the individual word level. This means that at least half of the actual words spoken are either not recognized, or are recognized incorrectly. It is an established fact that phoneme recognition is at least 10-20% less reliable than Speech-to-Text due to the small size of the recognition unit. In another well-known study\(^4\), eight leading speech research groups, including Cambridge University, AT&T, SRI, BBN, and John Hopkins University, participated in a National Institute of Standards (NIST) evaluation of telephony speech word recognition. Their results show significantly worse performance on the phoneme level as compared to the word level. Considering this compelling research, we can conclude that Speech-to-Phoneme reliability is at best around 30-40%. This means that two out of three phonemes are either not recognized, or are recognized incorrectly.

In contrast, UTOPY’s phrase-driven method is significantly more reliable than its Speech-to-Text conversion method and dramatically more reliable than the Speech-to-Phoneme method at the individual word level. These differences become even more dramatic at the level of business topics, because business concepts are generally verbalized in phrases.

Consider, for example, a phrase made up of three keywords, such as “didn’t receive (a) bill”. Let’s assume 50% reliability for each individual word, which is the best rate established by the above-mentioned studies for transcription methods. If we multiply the probability of correctly recognizing all three words together (50% x 50% x 50%), we find that the probability of correctly recognizing an entire three word phrase is only 12.5%, meaning that nearly nine out of ten times, the entire phrase will not be correctly recognized!

Results from Customer Tests
The fact that UTOPY’s phrase-driven approach is much more reliable than transcription methods has also been independently verified by the results of side-by-side tests that prospects have conducted between UTOPY and competitors who use transcription methods (Speech-to-Phoneme or Speech-to-Text).

In the example illustrated by the following chart, UTOPY and a competitor were both given the same set of calls and were asked to find specific statements made by call center agents or customers on those calls. The number of such occurrences properly recognized by each vendor is illustrated in the chart in Figure 1.

---

\(^3\) Min Tang, Bryan Pellom and Kadri Hacioglu, “Call-Type Classification and Unsupervised Training for the Call Center Domain”, University of Colorado at Boulder

\(^4\) Shuangyu Chang and Steven Greenberg, “Linguistic Dissection of Switchboard-corpus Automatic Speech Recognition Systems,” University of California, Berkeley
UTOPY’s Phrase-Driven Approach: Five to Nine Times More Reliable

This customer test and others undeniably confirmed that UTOPY’s phrase-driven approach is five to nine times more reliable than transcription approaches. For example, in the test illustrated in Figure 1 above, UTOPY correctly recognized more than nine times as many phrases related to the topic “No Prepayment Penalty”.

This fact was also confirmed by rigorous testing that UTOPY performed on the three different Speech Analytics methods. The results of these tests are shown in the following chart (Figure 2).
These tests confirmed that at low rates of detection, all three methods can be tuned to be highly accurate. However, as detection goes up, accuracy generally goes down. As illustrated by the chart, the rate of degradation in reliability at higher detection rates is dramatically worse for the Speech-to-Text and Phonetic transcription methods compared to UTOPY’s phrase-driven direct speech recognition and analysis approach. For example, at a detection rate of 70%, UTOPY’s phrase-driven approach is 82% accurate, while Speech-to-Text is only 11% accurate at that detection rate, and Phonetic is only 6% accurate at that detection rate.

The Importance of Completeness (High Detection Rates)

Agent Performance Optimization: Needs identification difficult with low detection rates
The impact of these differences is multiplied if Speech Analytics data is to be used for the purposes of critical business endeavors, such as agent performance optimization, which require at least two further levels of detail. For example, in order to improve First Call Resolution, we would need to find out what types of customer questions are prompting which agents to schedule callbacks. UTOPY precisely uncovers such granular correlations, and then automatically leverages this knowledge to recommend the best actions to improve business performance.

In the side-by-side test illustrated previously (in Figure 1), UTOPY correctly identified 149 callbacks that were scheduled, while the competition only correctly identified 37. In this case, five different agents handled this particular set of calls, and the calls were for three different call reasons. Each call reason had an average of five different sub-topics. Dividing the 149 calls with callbacks scheduled identified by UTOPY evenly between the five agents, three call reasons and five sub-topics, there would be two correctly identified calls per agent per sub-topic. If we make the same calculation on the 37 calls UTOPY’s competitor correctly identified, the result is .5 --- ½ of a call per agent per sub-topic correctly identified on average by the transcription method. Obviously, it is much easier to identify what types of questions are prompting which agents to schedule callbacks with two calls per agent per sub-topic.
compared to ½ a call per agent per sub-topic. UTOPY leverages this more reliable data to automatically identify which customer issue(s) each agent has difficulty resolving on the first call from the customer, and UTOPY automatically recommends the best action(s) to take to improve each agent’s knowledge of the particular topic(s), recommending, for example, specific coaching, training or e-Learning for each particular sub-topic.

Measuring the Impact of Training, Coaching or Policy Changes: Time is Money
When changes are made to training/coaching programs or business policies in order to improve contact center performance and the customer experience, UTOPY immediately measures and continuously monitors the impact of such changes. Because of the high level of Speech Analytics data reliability it delivers, UTOPY can measure the impact of such changes in much less time with much less data than the competition. Since their data reliability is five to nine times lower than UTOPY, competitors require five to nine times more calls (data) to analyze in order to uncover the same information. Much more time is required for a typical contact center to make or receive so many more calls.

Decision Making and User Adoption: Decision-makers can’t rely on anecdotal data
Decision-makers shouldn’t rely on analytics or reports if they are based on data that is not complete. If the data is not reliable, people will not trust or use any calculations, reports or other analysis that is based upon that data.

Compliance: The Impact of Missing “Non-Compliant” Events
Adherence to regulatory or enterprise policies and procedures is a prerequisite to being in business in many industries today. The majority of a company’s interactions with its customers will comply perfectly with such regulations. However, the occurrence of even a single non-compliant event, no matter how rare such an occurrence is, could cause extensive litigation and other trouble and expense for a company. It is mission-critical in such situations to detect and correct every single occurrence of a “non-compliant” event.

Customer Satisfaction (CSAT) and Loyalty: The ability to maximize CSAT increases with the amount of customer sentiment detected
Customer satisfaction can be difficult to measure, yet any increase in satisfaction means a corresponding increase in customer loyalty and up-sell/cross-sell potential. In order to increase customer satisfaction, all available customer sentiment must first be reliably detected and measured. Any customer dissatisfaction must be detected, the reasons for that dissatisfaction must be completely and accurately uncovered, and then proactive action must be taken to improve satisfaction levels. For example, each unhappy customer should be identified and proactively contacted in an attempt to keep that customer’s business.

The Importance of Context: Phrases provide context to the spoken word
A single word can be used in a myriad of different ways, with many different meanings. For example, the word “bill” could refer to: a person named “Bill”; or could be part of a request to speak to the “Billing” Department; or could be someone saying that they didn’t get a “bill”; or that he/she refuses to pay a “bill”. These differences in meaning are very important, yet only the entire phrase provides the
context which enables true understanding of the real meaning of the word. Only UTOPY’s phrase-based methodology delivers the true meaning of speech.

**Conclusion**
Reliable Speech Analytics data, which is accurate, complete and delivered in its business context, is the prerequisite to optimizing contact center performance and the customer experience. UTOPY’s highly reliable data foundation enables targeted business applications which produce actionable information, driving best practice workflows for optimizing business performance, and propelling UTOPY towards the opposite end of the business value continuum from the rest of the Speech Analytics vendors, whose technology is exclusively transcription based.