HGS® digital

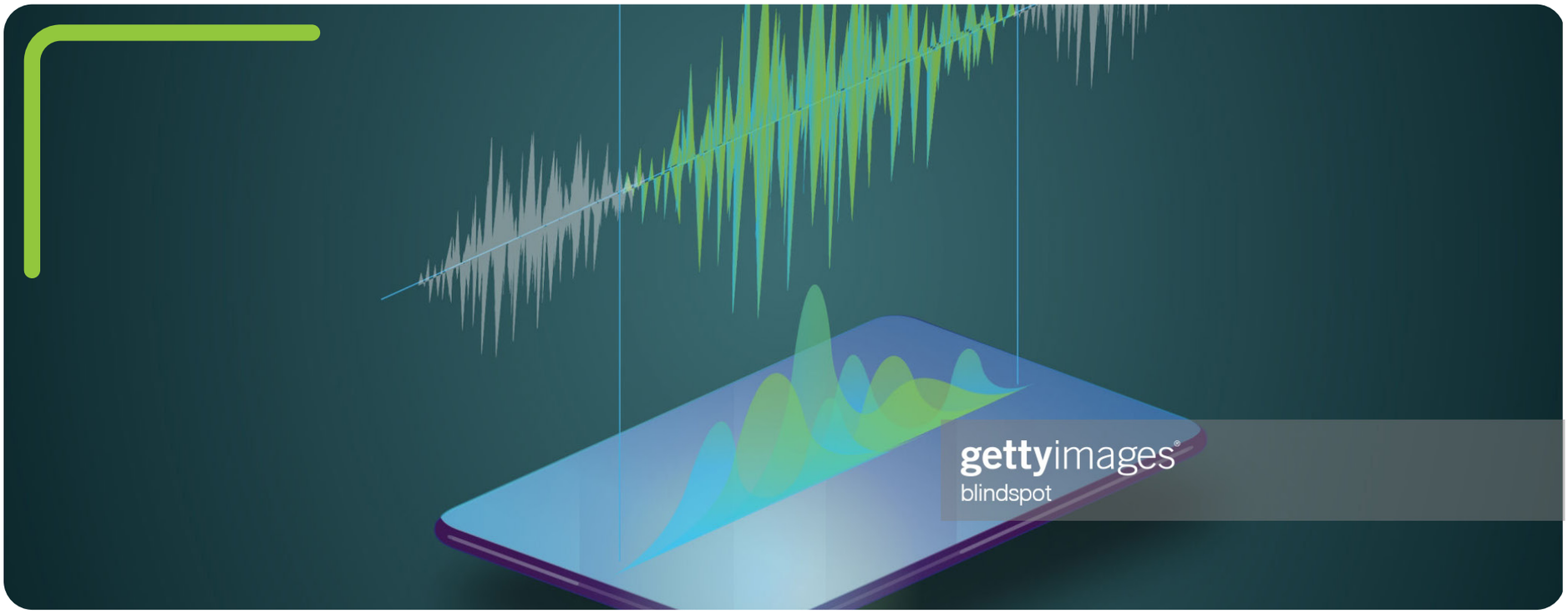
SPEECH ANALYTICS STEERS REVIVAL OF TRUE VOICE OF THE CONSUMER, WITH MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE SUPPORT

CALL IMPACT OF TODAY'S DISRUPTERS: FROM ENHANCED CSAT TO ACTIONABLE INSIGHTS

Introduction

We are in the midst of a resurgence in terms of the role voice can play in delivery of an optimized customer experience (CX). A lot of the credit for that goes to the capabilities and applications of today's speech analytics, which increasingly show dramatic ROI as a result of the evolution of disrupters like AI, machine learning, neural networks, automation, NLP, text mining, and signal processing. According to Gartner, phone calls are still an important channel for research and purchasing, with 70% of mobile searchers reporting they click-to-call directly from the search results to connect with a business. And recently, BIA/Kelsey estimated that this year consumer-to-business phone calls made from smartphones is a market that will grow to nearly $170 billion.

Yet, businesses are not responding in alignment with consumer need. Our own research shows there are multiple factors that contribute to the failure of businesses to optimize analytics—from lack of bandwidth and skills to a resistant culture. In fact, 50% of businesses surveyed by HGS on the topic of analytics "do not have the resources to help deliver what we aim to," while the other half "have different opinions on analytics" and also "do not have the resources to help deliver what we aim to."



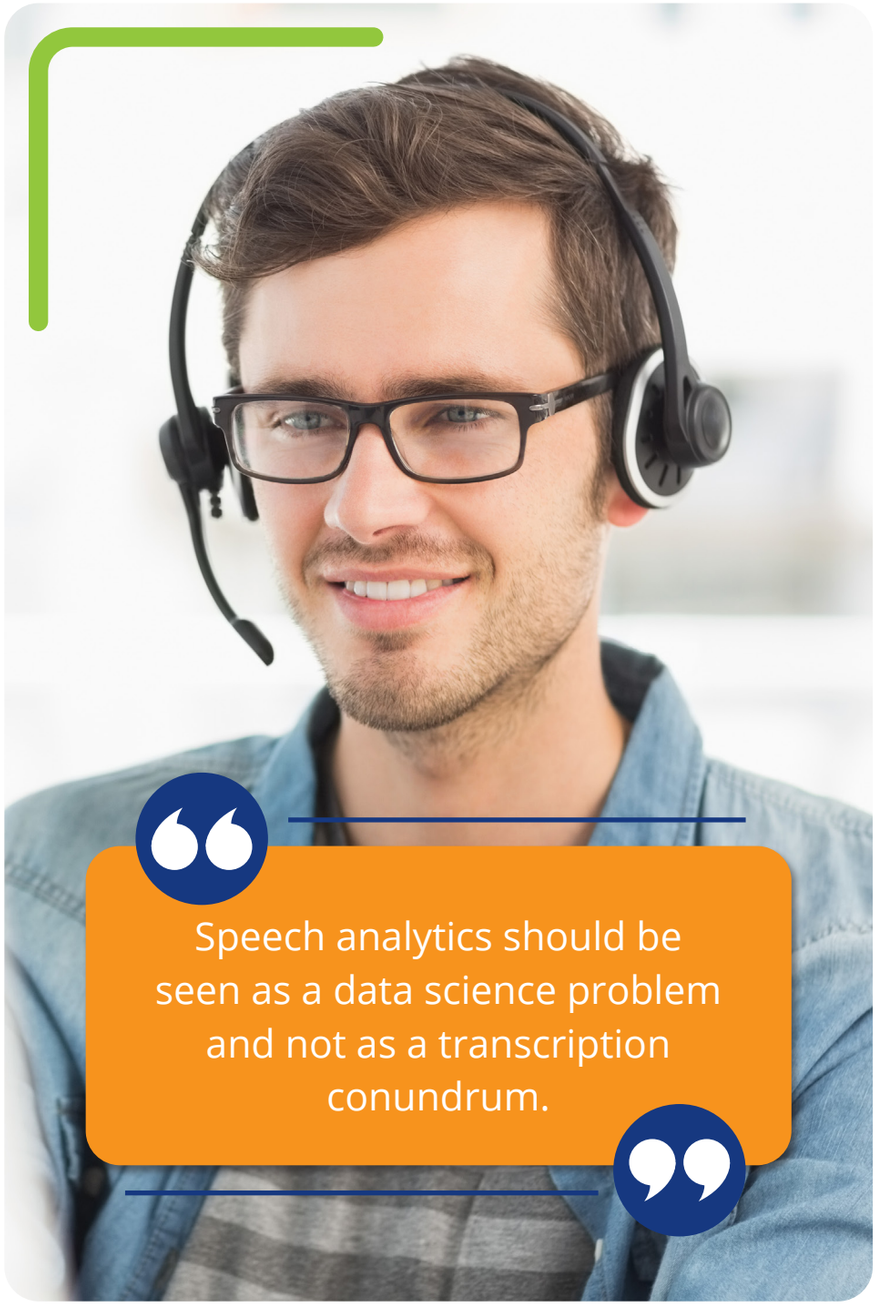
The Evolution of Speech Analytics

When call center personnel hear conversations across geographies, they tend to hear the sounds they are accustomed to. In the wake of an onslaught of accents, dialects, conventional slangs, lexicons, and cultural references, the context of the calls is marginally lost. Add to the mixture a labyrinth of problems stemming from local issues, including: bad cell phone coverages, signal overlaps and interferences, and speakers' emotional states, and, it is almost as if the entire premise of making a phone call and talking to a human to resolve an issue is futile.

To mine for the action items of these conversations, speech analysis is a comprehensive tool aimed at optimizing issue resolution by taking actions on text deciphered from the communication.

Acting upon speech is very subjective, quality constrained and is speaker-agent dependent, while machine automated text of the speech gives us measurable action items under quantifiable metrics. This ensures higher quality control and more satisfied customers.

> While we listen to the caller, the system deciphers facets of the call to act upon.



> Speech analytics should be seen as a data science problem and not as a transcription conundrum.

Speech Analytics Framework and Approach

Let us start with the ideal speech analytics model:

Agent Performance Analysis – Overall Solution

Data Source	Data Preprocessing	Feature Extraction	Model Building	Reporting
Call Recordings	Analog to Digital conversion of audio signal	Derived variable for Deep Learning Modelling	Speech-to-Text Conversion Model using Deep Learning	Agent Performance Evaluation
Audiobooks	Noise Reduction	MFCC, Spectrogram, RASTA, Cochleogram	NLP on Text Data to Evaluate Agent Performance Quality Score	KPI Dashboard
	Improve Signal to Noise Ratio			

Speech analytics, within the framework of commercial usage and its applications in the call center paradigm, is mainly centered around the speech-to-text conversion as measured by the automatic speech recognition (ASR) metric. The biggest flaw in the industry of analytics and automation lies here. There is not enough heed paid to the quality of the audio file itself or post-transcription word post-processing. Quality checks are all manual in nature and rarely computer-aided solutions are considered. The ideal flow of speech analytics work should be as:

SIGNAL PROCESSING → **TRANSCRIPTION** → **SPEECH MINING AND TEXT ANALYTICS**

To start with, customers/callers usually hear the statement "calls may be monitored or recorded for quality and training purposes" and that includes the condition that the second level analysis of calls and responses is also manual and hence subject to human error.

These issues can be mitigated by a thorough development of knowledge base in these areas:

> Speech analytics is a three-step process. Speech is a signal before it becomes an audio file and is a set of characters after it becomes text.

Term	Description	Components	Metrics
Speech Analysis	Study of speech sounds for purposes other than linguistic	• Speech recognition • Medical analysis of the voice (phoniatrics) • Speaker identification • Emotional state of speakers (voice stress analysis)	• ASR • Voice analysis • Accent and Stress analysis
Speech Processing	Aspects of speech processing include the acquisition, manipulation, storage, transfer, modeling and output of speech signals. The input is called speech recognition and the output is called speech synthesis.	• Digital Signal Processing • Stochastic Signal Processing	• SNR • PNR • Gap/Silence Analysis • Interference
Speech Recognition	Computational linguistics program that develops methodologies and technologies which enables the recognition and translation of spoken language into text by computers.	• Transcription • Text Cleaning	• Standard text mining metrics • Word error rates
Speech Synthesis	Artificial production of human speech or text to speech	NA	NA
Speech Analytics	Process of analyzing recorded calls to gather customer information to improve communication and future interaction	• Text Analytics and Mining – Topic Modeling (LSA, LDA) – Sentiment Analysis	• Call center legend • Training QA & QC • met1rics

Relevance to My Business

Most companies, irrespective of their revenue leaning, do tend to support call-based consumer service and customer care centers. However, they need to up the ante on speech analytics to solicit enough information from the call to make it actionable without escalations. This can only be achieved by recognizing speech as a collection of cleaned and usable textual characters; characters that were developed from signal waveforms, characters that contain hidden information and myriad patterns that need to be unlocked and unfolded.

The tangible business milestones for archival in this process include:

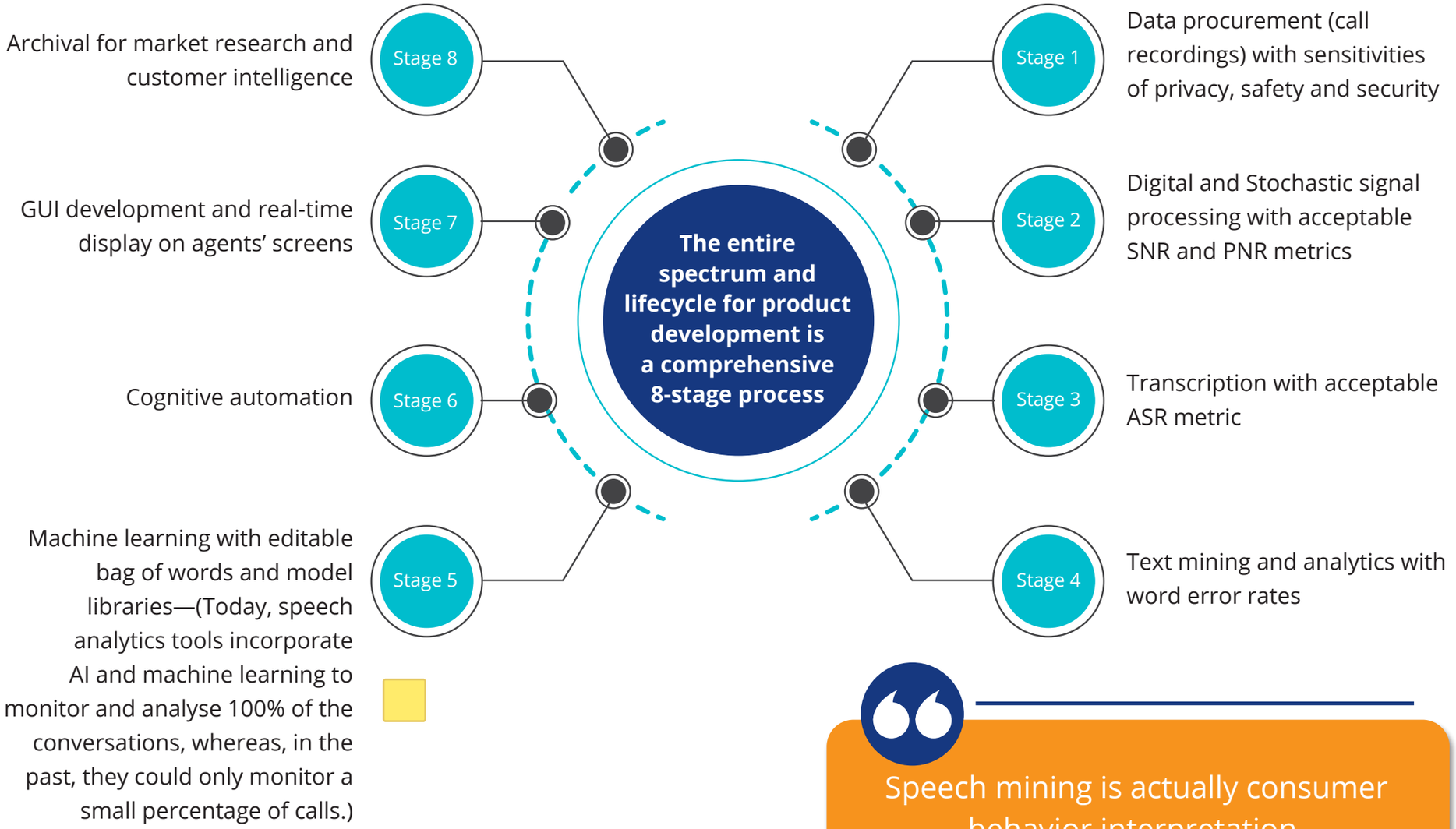
- Monitor support agents and representatives.
  - Ensure Process Automation: Calculation of Agent Performance Matrix.
- Ensure emotion and sentiment analysis of customer call.
- Identify case types or problem types.
- Pinpoint the issues for which people need the most help.
- Conduct complaint analysis.
- Ensure call (and caller) categorization.
  - Ensure training and quality control.
- Protect user data privacy.
- Optimize costs (labor, training and material).
- Create your own intellectual property.
- Ensure you have trainable models that focus on learning.

Our recommended proposed speech analytics approach is:

Module	Explanation	Activities	Sub-Activities
Audio Separator Engine	Speaker ID Identification	Unique speaker modeling based on wavelength, amplitude, bass and frequency analysis	Accents
Voice Analysis	Emotional state of the speaker	Tone, Volume, Tenor, Pitch, etc.	Intonation, Stress, Rhythms and Disfluencies
Audio Cleansing	Acoustical signal to digital signal	Signal transformations, SNR + PNR and Spectral analysis	Signal cleansing (Noise reduction)
Speech Recognition Engine	Convert Digital Signal to Phonemes, then words	Transcription	Attach ID to each word
Text Data Cleansing	Standard checks for making textual data usable	Conversion to lower case, removal of stop gaps, stop words, punctuation marks, numbers, special characters, word trimming and word stemming	Auto-cleaner developed
Audio Sentiment and Topic assignment for each ID	Extract sentiment, topic, sub-topic for each work	LSA, LDA and Probabilistic sentiment analysis	Context, meaning, relevance, purpose and intent to be experimented with
Text Mining	Word Associations	Extract correlations, priority and triage of each topic to define case types and action items	
Deep learning algo for performance matrix	Develop a self-learning, autonomous and ready-to-deploy expert system that automates all these processes	NN based cognitive self-learning models	Error minimization, big data support and model optimization
Front End	Simple GUI to play around work with I/O		
QA & QC	Testing our models with accuracies seen across third-party solutions	• ASR • Word error rate • Data cleansing	• IBM • Google • Microsoft, etc.
Testing	Testing models developed across various sample data sets	• Training data with mock numbers • Multisector voice samples	• Aviation • Healthcare • Telco • BFSI

Getting Started with Speech Analytics

A ground-up approach is recommended for all companies in the space.



> Speech mining is actually consumer behavior interpretation.

Benefits of Speech Analytics

From understanding how people speak in a context to interpreting consumer anger in numbers, there is significant speech analytics ROI for today's businesses. Some concrete business numbers would be seen in:

- Improved average handle time (AHT), turnaround time (TAT), quality assurance, and control and CSAT metrics
- Increased conflict resolution
- Reduced consumer complaints
- Improved situational awareness of agents and executives
- Improved vocabulary and information retrieval of agents and executives
- Improved training and QC methods
- Improved rates of cross-sell and up-sell
- Targeted and actionable intel from suggestions and feedback through VOC
- Manpower optimization in support teams (QA/QC)
- Increased accuracy in evaluation due to better sampling
- Pointed feedback and improvement plan

> Customer segmentation, targeted marketing, and consumer profiling are a direct consequence of the process.

Looking ahead, speech mining and automation will work together for:

- Real-time transcription of legal courtroom proceedings
- Real-time transcription of interviews of potential candidates during hiring
- Real-time transcription of interviews on TV and subtitling
- Forensic voice analysis
- Real-time transcription of healthcare practitioner and patient conference in telemedicine
- Real-time transcription of teacher-student interactions in tele-education
- Real-time transcription of remote meetings through teleconferencing in an office room setting

Common speech mining and analytics-centered activities include:

- Account Management
- Sales
- Partner Management
- Marketing and Advertising Marketing and Advertising

Ramp and Cost

Our recommended ramp based on 8-stage developmental goals will typically scale in 90-180 days, in phases:

1. **Digitization** – Converting analog signal into digital representation
2. **Signal processing** – Separating usable speech from background noise
3. **Phonetics** – Understanding variability in human speech
   - Phonology – Recognizing individual sound distinctions (similar phonemes)
   - Lexicology and syntax
     - Disambiguating homophones
     - Features of continuous speech
     - Syntax and pragmatics: Interpreting prosodic features
   - Pragmatics: Filtering of performance errors (speech disfluencies and speech impediments)

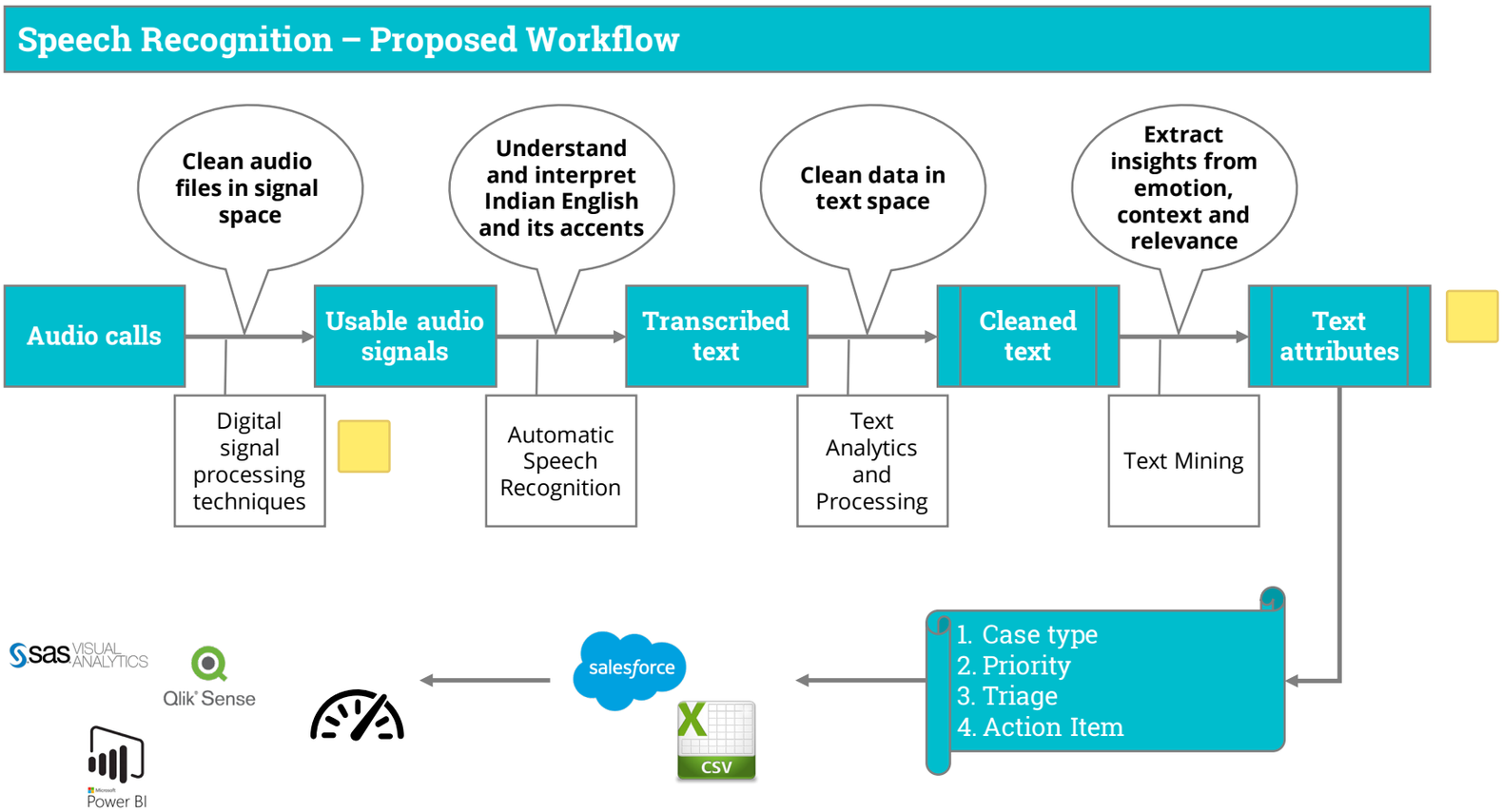
From a non-computational linguistics appeal, there is a gamut of complexities that need to be overcome, to fathom any sort of mastery on speech analytics in textual communication:

Word confusability	Ellipses and Anaphors	Multiple Interpretations
Word variations	Limited vocabulary	
Speaker Dependency	Context-dependency	Coarticulation
Noise/reverberation	Intra-speaker variability	

Budgeting and costing in this regard, are provided by module rather than product based. The product life-cycle development should ideally be proportioned into five categories:

- Signal Engineering
- Transcription
- Speech Mining
- Automated Information Display
- Cognitive Intelligence Generation

Our speech recognition proposed workflow is a focus on audio to action item:



With strong agent performance tracking in speech analytics, reporting is standardized and benchmarked for future success:

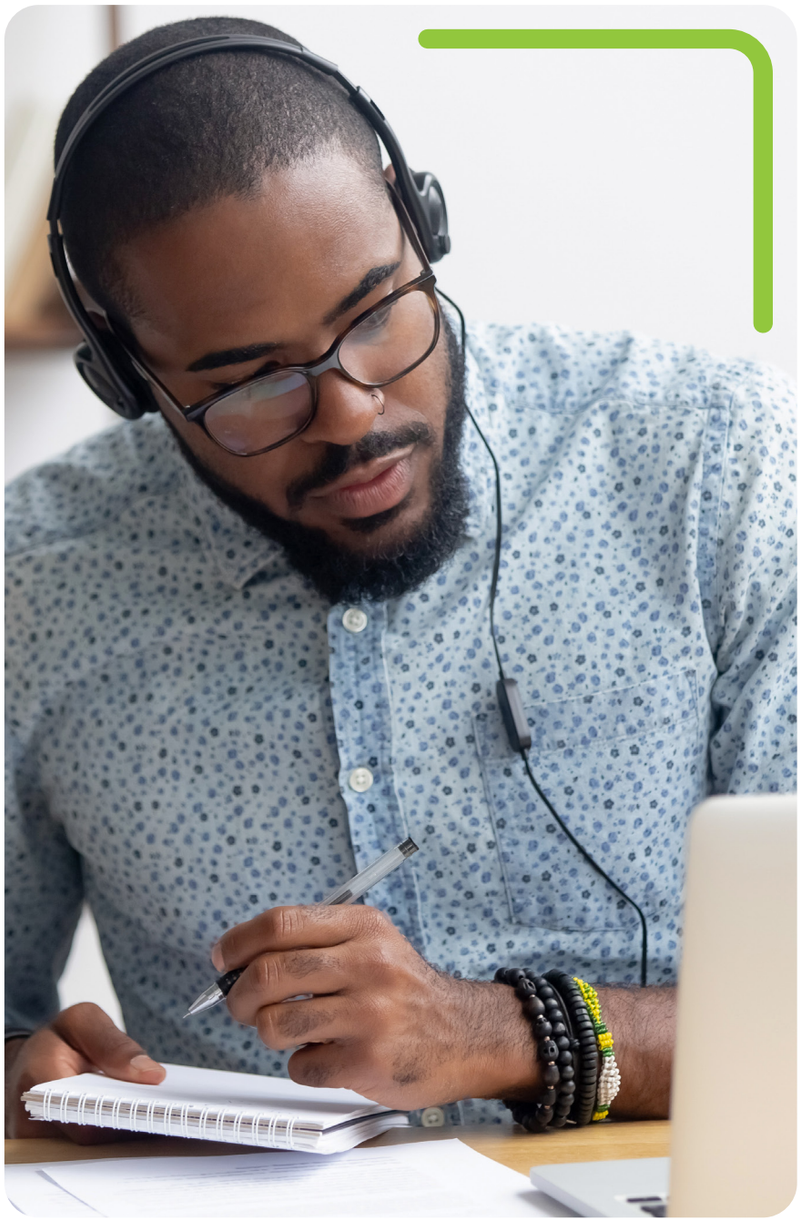
Individual categories should be listed for module-based line-unit development as summarized below. Costing would then happen based on choice of hybrid R&D model, or complete in-house development, or vendor-based solution management.

How Service Vendors Can Help/BPO Selection

When assessing return on speech analytics investment, businesses will experience results across stages. It is imperative to note that each of the three speech analytics focus areas (Signal Engineering, Transcription, and Speech Mining) can be broadcast as stand-alone business ventures and solutions. Consequently, a concurrent omni-role set up thus be allowed for cost-center profit-center definitions.

- The first signs of positive ROI would appear in the signal engineering space, where input audio/speech processing would reduce the size of the raw files by filtering out silence, background noises, corruptions and speech disfluencies. Lossless compressions would provide savings in processing times, computational expenses and storage spaces.
- The second tangible benefit would be seen in the manual quality control (QC) process where the quality teams now listen to streamlined audio samples rather than the original corrupted files. Although this is the process intended to be automated, the manual process itself will not be replaced.

Ultimately, the strongest results will relate to the and corresponding metrics and computational times. The higher the ASR and the lower the computational times, the greater the revenue generation and cost savings. Incorporation of transliteration and translation services as value-adds brings multilingual support and associated challenges, risks, rising costs, and stringent demands but equivalent gains.



Another revenue stream would stem from the text mining space, where measurable intelligence will be generated through simple ac items for agents to comprehend and sort to; an otherwise humungous reading will be minimized to an optimised to-do list with a do's and don'ts checklist requiring one-step mouse clicks.

Today's automation would allow for improved response times, reduced TATs, decreased customer conflicts, lowered AHTs, lessened customer escalations, increased case resolutions and thus augmented CX levels and CSAT scores. It has been noted that CSAT scores share a direct relationship with PSAT (partner) and ESAT (employee) levels as well, meaning gained CSAT trust would have benefits that proliferate and trickle down and across.

This will lead to strengthened engagement levels with partners and elevated employee-employer relationships. Agents, QC experts, managers, and trainers already operating in a volatile, sensitive and high-stress environment will tend to relax enough to boost their occupational health and psychometric evaluations.

Parallel and independent revenue streams can also be opened up in the two remaining focus areas of the study (Automated Information Display and Cognitive Intelligence Generation) which deal extensively with dynamic text based solutioning, real-time displays and, if required, client side/local business deployment.

Smart organizations will also use valuable insights gained from customer data mining to generate cross-revenue from market research, consumer segmentation, customer outreach, upsell and cross-sell and loyalty programs.

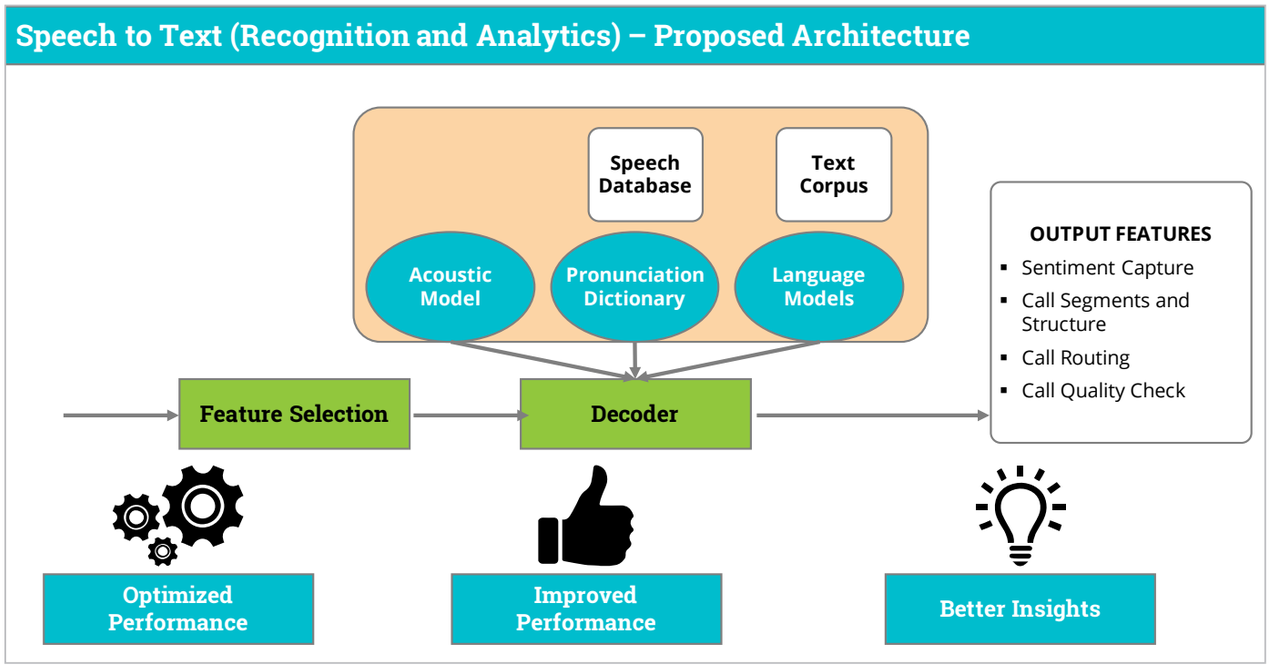
HGS Digital Distinctiveness

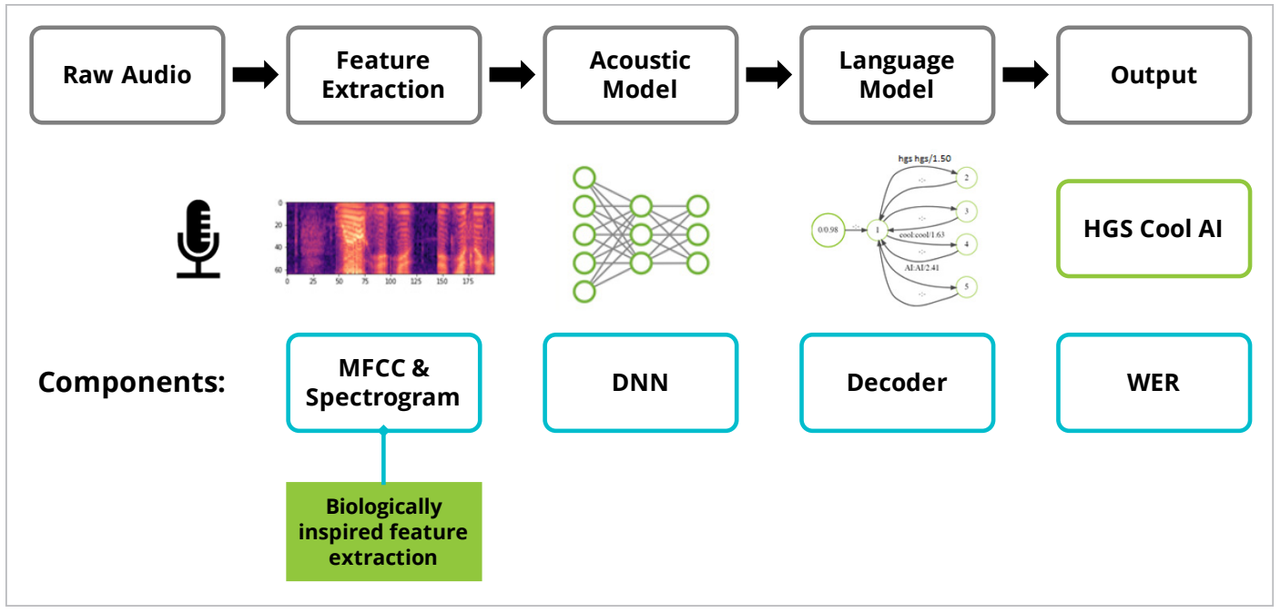
HGS Digital has more than 40 years of experience incubating innovation and launching proof of concepts (PoCs) for our clients. For speech analytics, we've built a resource pool of people, processes, and technology innovations. Here is one use case comprising a standard research and development process:

Stage I of the R&D process, with use case structure:

Speech Analysis Application	Description	Component Use
What are customers escalating?	"I WANT TO SPEAK TO YOUR SUPERVISOR!"—the words that no contact center agent ever wants to hear. Escalations are the symptom of something terribly gone wrong or customers' perceived injustice. Lastly, escalations unchecked can lead to a more serious type of speech analytics query, which includes mentions such as: • Attorney general • Better Business Bureau • My lawyer (attorney) • See you in court! • Consumer Financial Protection Board (CFPB) • File a complaint	Speech Recognition Engine
Pulling calls for coaching vs. pushing moments that demand attention	Random coaching practices are prevalent and not purposeful across the contact center industry today. Speech analytics should be used to target coachable interactions. OpenText uses evaluation plans which are custom-filtered evaluation play that set: • Topics based on interactions tagged by speech analytics, surveys or metadata • The frequency of coaching for that agent group • The distribution of coaching work across a QM team, team leaders or coaches • Injected calibration interactions that insure consistency of coaching	Topic Modelling (LSA, LDA)
First Contact Resolution (FCR) improvements	FCR is a longstanding contact center metric that is often plagued with calculation errors, disagreements and confusion. Regardless of the calculation method, organizations do monitor repeat callers for a reason—customers do not like to call repeatedly and not handling the issue on the first contact comes at a cost. Speech analytics takes the guessing out of FCR by searching for key phrases such as: • Second time I've called	Sentiment Analysis

Stage II of the R&D process elaborated design and development of corresponding solution architecture and workflow:





Stage III of the R&D process required design and development of corresponding outputs with error metrics and computational times:

Actual Transcript	ASR Model Output	Error
Since then physicists have found that it is not reflection, but refraction by the raindrops which causes the rainbows.	since then physicists have found thought reflection or fraction by the range of which causes the re emergence	0.526316
No details of the talks were given.	no details of the talks would give in	0.428571
Clearly, the need for a personal loan is written in the stars.	clearly the need for buffalo in his versus south	0.666667
Sadly, the revival could not be sustained.	sadly the revival Kenobi sustained	0.428571
The confidence is low, but it is a difficult thing to understand.	the confidences let it is difficult thing to and found	0.583333

Error Measure:

$$\textbf{Word Error Rate} = \frac{\textbf{Insertions + Deletions + Substitutions}}{\textbf{Number of Words in Reference Transcript}}$$

Stage IV of the R&D process invoked the use of modern text mining exercises to extract meaning, relevance, context, mood, intent, and purpose from transcribed text:

Role of Preprocessing:

- Removal of stopgap words, general words and speech disfluencies
- Remove unwanted numbers, special characters and grammar points
- Stem words to their English root
- Removal of white spaces and artifacts of transcription
- Conversion of all words to lower case
- Creation of a document term matrix

Role of Sentiment Analysis:

- Based on customer escalation, assign probabilistic sentiment to calls
- Create and dynamically update scores/indices for agents based on escalations
- Identify emotion of the caller based on tone and tenor of the caller

Role of Topic Modeling:

- Identify type of calls (complaint, cross-selling, up-selling, etc.)
- Identify topics and subtopics of calls
- Based on topic of call, assign priority and triage (urgency) to call

Role of Word Associations:

- Wordclouds – most frequently appearing words (topics)
  - Intent, Meaning and Purpose
- Heat Maps
  - Frequencies and Correlations
- N-grams – most frequently appearing chain of words (word links)
  - Context and Relevance

Role of Scoring:

- Identify right agent to respond to the call based on a unique performance-centric pattern matching algorithm
- Identify and designate priority
- Assign triage (urgency)
- Provide bag of words indicating keywords and key phrases to be part of the conversation to make the overall call vibe positive

Stage V of the R&D process required the use of in-house developed tools (and hence the innovation) to dynamically refresh Salesforce screens of agents and Excel dumps of QC personnel and marketing team, with processed information seen in discernible columns, appended to the original calls as additional attributes.

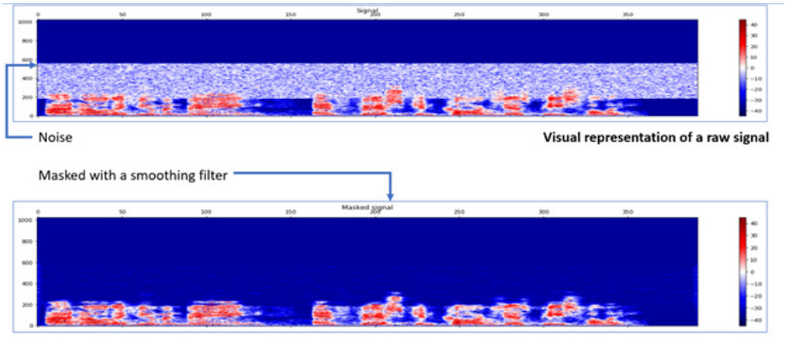
The following tables briefly summarize the algorithms and tools used in this process:

	Technique	Tool	Method
1	Noise Reduction	Python	FFT and Spectrogram Analysis
2	Silence Detection and Removal	Python	Signal Energy and Spectral Centroid, Thresholding
3	Audio Feature Extraction	Python	Spectrogram, MFCC
4	Acoustic Modelling	Python	Deep Learning: RNN-LSTM
5	Decoder	Python	CTC Decoder
6	Language Modeling	Python	n-gram, Kneser–Ney Smoothing
7	Topic Modelling	Python	LDA, LSA
8	Sentiment Analysis	Python	NRC, Syuzhet, BlobText
9	Word Associations	R/Python	tm, SnowballC, ggplot2 and Wordcloud

Sample outputs are seen below for use and reference:

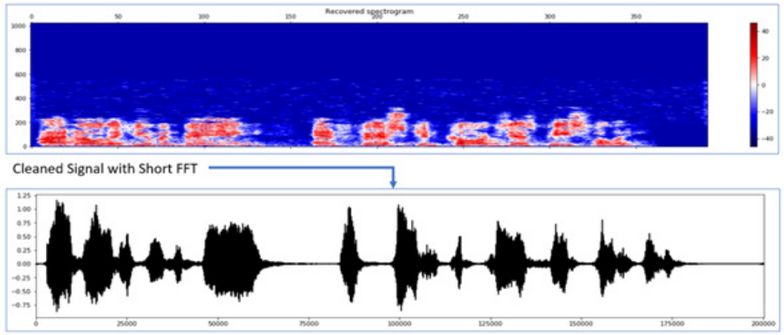
Output 1 (generated)

Audio Pre-processing (Digital)



Output 2 (generated)

Audio Pre-processing (Digital)



Output 3 (researched)

Audio Pre-processing (Stochastic)

Modeling

- Studying Stationarity, Ergodicity, Spectral distribution and density of spectrogram
- AR, MA and ARMA models for Spectrogram analysis

Estimation and Correlation analysis

Parametric estimation

Non-Parametric Spectral analysis

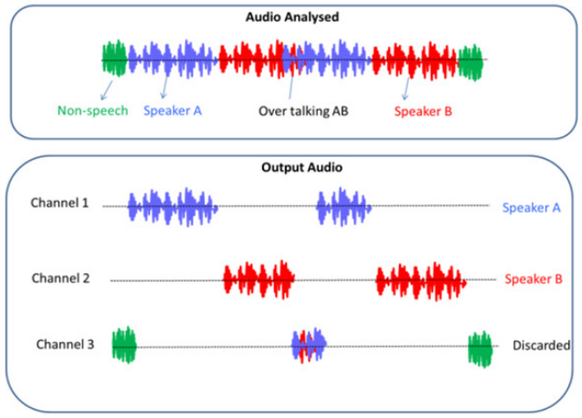
Correlation matrix (comparing Signal with Noise for additional smoothening of Signal)

Output 4 (emulated)

Speaker Identification and Diarization

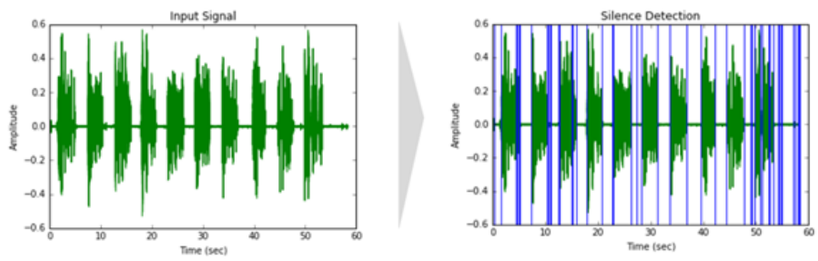
Process of partitioning an input audio stream into homogeneous segments according to the **speaker** identity.

- Wikipedia



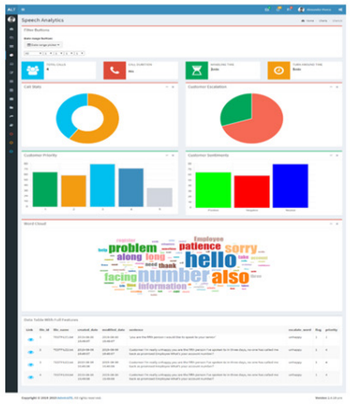
Output 5 (emulated)

Silence Detection



Output 6 (generated)

Dashboard - KPI: Manager view

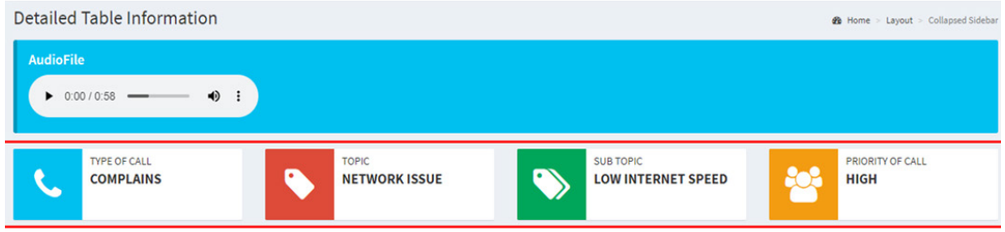


Dashboard - KPI: Customer Escalation

Data Table With Full Features

Link	file_id	file_name	created_date	modified_date	sentence	escalate_word	flag	priority
	1	TESTFILE1.txt	2019-08-06 16:49:07	2019-08-06 16:49:07	'you are the fifth person i would like to speak to your senior'	unhappy	1	1
	2	TESTFILE2.txt	2019-08-06 16:49:07	2019-08-06 16:49:07	Customer I'm really unhappy you are the fifth person I've spoken to in three days, no-one has called me back as promised Employee What's your account number?	unhappy	1	4

Dashboard - KPI: Topic Modeling



Dashboard - KPI: Sentiment Analysis

The Way Forward – What's Next for Speech Analytics

Our proposed go-to-market approach is to deploy this solution at multiple in-house support centers in India and source a periodic feedback for quality and training purposes:

- HGS Digital provides call-based customer support at multiple locations across the world for various industry sectors.
- Employ these contact centers as launch pads for technology acceptance and endorsement.
- Collect UAT reports from them, specifying metrics for business approval and validation, including:
  - Ease of use and user-friendliness
  - Reduction in computational times
  - Reduction in resolution times
  - Reduction in human errors
- Source as many archived conversations as possible to train and refine our models to better our accuracy metrics.
  - The labelled data would come from our on-going endeavor to manually transcribe calls to build a robust 100 – 1,000-hour word bank across industry verticals and domains

Why HGS Digital?

Many large-cap firms and contact center organizations are consciously collecting massive amounts of demographic, transactional, ancillary and collateral data from across cooperative omni-channels, subject to regulatory frameworks and legal obligations. HGS Digital has been in the forefront of such an endeavor. By investing in our Data Science and Analytics Program, HGS Digital has built a diverse portfolio of customized solution development and deployment, across multiple industry verticals and domains:

- Media
- BFSI
- FMCG
- CPG
- Telecom
- Healthcare
- Pharma
- E-Commerce
- Automotive

- HR Analytics
- Retail Analytics
- Marketing Analytics
- Operational Analytics
- Sales Analytics
- Content and Viewership Analytics
- Competition Analytics
- Telecom Analytics
- Social Media Analytics
- Financial Analytics
- Geoanalytics

HGS Digital Analytics and Data Sciences Services developed Speech Mining solution stands-out, with abilities in:

- Thorough speech audio cleansing and preprocessing in the signal space
- Quality speech recognition and transcription unique to Indian English with accuracy comparable to other COTS
- Contact center QC legends based learning and scoring systems
- Assured speech mining and text analytics based actionable intelligence
  - Probabilistic (vs. definitive) emotion, topic, subtopic, sentiment and case-type assignment
  - Scoring based priority and triage bucketization
  - In-house pattern matching algorithm for agent assignment
  - Recommendation engine to suggest a bag of keywords, to be used as part of a response to ensure overall conversation vibe to be at least 50% positive
- Seamless on-site deployment
  - Real-time dashboarding with dynamic refreshing
  - Accurate isolation and labelling of discernible patterns
  - Robust anomaly and outlier detections
- Providing value additions via projecting future trends and forecasting business metrics
- Developing and listing editable, trainable and customizable libraries relevant to the Indian cultural context
  - Unhindered access to statistical models driven by in-house designs
  - Fine-tuning weights and parameters of various ML libraries based on automatic error minimization
  - Unique model designs customized to various industries while being largely domain agnostic
- Comprehensive testing on 100s of multi-sector calls
- Being computational optimized and cost moderate
- Operationally friendly based on feedback received from the floor
- Periodic updating and upgrading keeping abreast of latest advances in ML and AI technologies

HGS DIGITAL

With over 250 digital successes and stellar client satisfaction ratings, HGS Digital creates frictionless customer experiences that solve complex business problems and improve people's lives.

We work with leading brands across the world to improve their customer engagement, optimize their operations, reduce costs, and increase revenue. As a technology-agnostic consultant—with partners ranging from Amazon, Salesforce, Google, IBM, Microsoft, and more—we are well-equipped to help you select and successfully implement the right tools for your specific needs.

www.hgsdigital.com

sales@hgsdigital.com