

A REVIEW PAPER ON VOICE ANALYTICS

¹Charu Mishra , ²Professor Aarti Mehta Sharma

¹MBA 2nd year student, Symbiosis Institute of Business Management, Bengaluru, A constituent of Symbiosis International University, (India)

²Faculty Guide at Symbiosis Institute of Business Management, Bengaluru, A constituent of Symbiosis International University, (India)

ABSTRACT

The document presents a literature review on Voice analytics and briefs about the various U.S patented inventions and the technologies that are used to perform voice analytics. Apart from this, it describes the application of voice analytics in various fields and gives a brief about the various players involved in voice analytics market and how it has grown over the years. This shows that voice analytics has huge potential and efforts are needed to develop it further so that new technologies can be developed and integrated with other fields.

Keywords: Application, Business, Inventions, Potential, Voice Analytics

I INTRODUCTION

With the advent of technology, the amount of data keeps piling up making it difficult to analyze huge amount of data to give some meaningful information. Many analytical techniques and tools have been developed by organizations to analyze and obtain business insights from data. Voice analytics is one such branch of analytics that helps various organizations in this purpose by analyzing spoken words or speech between two or more people. It is one of the fastest growing technologies because it helps in analyzing the sentiments of people during the call which can be useful in identifying their inclination towards a product or service. In some cases it also helps in detecting fraud. This paper will focus on the various processes which are used in voice analytics, its application in the real world with main focus on its use in the business world and what technology are used in performing voice analytics. The below figure depicts the basic process followed in any voice analytics technique:



Fig1: Describing the Basic Process Flow in Voice Analytics.

II LITERATURE REVIEW

Voice analytics is the branch of data science that deals with computerized processing of a (digitized) speech audio signal to extract information. The extracted information has various applications in the business world and can be used to make important business decisions. The most widely used application of voice analytics is in call center, where it is used to obtain useful information from both structured and unstructured calls between a caller and an agent. Apart from call center, its other application is in fields like bioacoustics, diagnosis of speech related disorder, financial markets, etc. On understanding its importance, a lot of research has been done in this field for more than a decade, and various technologies and methods are invented to perform voice analytics.

2.1 The Techniques involved in Voice Analytics

Over the years, scientists have developed many unique and innovative technologies and processes to carry out such a complicated but extremely important analysis. In the existing method, the supervisor randomly selects a call to detect a problem, but the problem with this method is that the call selected by supervisor may not be the one having problems and hence the chances of selecting a wrong call is high. The inventions made by various scientists described below are more systematic and are based on automatically detecting problematic calls that reduces the disadvantages of the current system. The various techniques described are as follows:

2.1.1 The process of Chan et al invention

[1] There can be two types of customers contacting the call center; an existing customer or a new customer. The deviation or problem in the calls of existing customers is identified by call center agents by comparing the calls with the existing profiles of the customers (that gives comparison criteria) in the database. If some deviation occurs from these criteria then the call is identified as problematic and is forwarded to the supervisor. In case of a new or unknown customer, normal profiles are used to determine the monitoring criteria.

Monitoring is also done by processing speech associated with the call and by identifying indicators of profanity, inappropriate language or the use of word 'supervisor'. If any of these indicators are found in the processed speech, then monitoring profile is searched to determine whether the speech is normal for the customer or the call should be directed to supervisor. Pitch and intensity of the call is also measured by detecting the energy levels of the call across a voice spectrum, in order to identify any abnormality. Call is diverted to the supervisor if an abnormality is detected. At the end of the interaction, these calls are saved and stored and saved in database so that the supervisor can play them again and determine criteria for call monitoring. This method described by Chan et al is mostly context dependant.

[2] Applebaum et al (2003) came up with an invention that was able to handle context independent calls. which was earlier not possible. Applebaum et al described the use of semantic categorization system, parsing system and Automatic Speech Recognition (ASR) system to solve this problem. ASR is a computer based method of identifying words spoken by individual in a microphone or telephone and converting the speech into written text form for further analysis. The aim of ASR is to allow the computer to identify all the words spoken by an individual intelligibly during real time call scenarios.

2.1.2 The process of Applebaum et al:

As the call center agent receive a call, the unstructured voice data obtained from the caller is converted into unstructured text data using ASR. As the call is unstructured, it becomes imperative to categorize this unstructured text data processed by ASR to ensure the purpose of the call which will determine the course of actions to be taken to handle the call. Such categorization is done by semantic categorization system which categorizes words based on the neural response produced by words stimuli in the brain region that are activated by such stimuli. Semantic categorization system uses a global parser that parses the input text data into semantic categories for classifying calls. Some of these categories can be complaint, product query, store location, etc. Global parser relies on local parser that provides syntactic tags (like phone number, address, Product-ID, etc) to keywords.

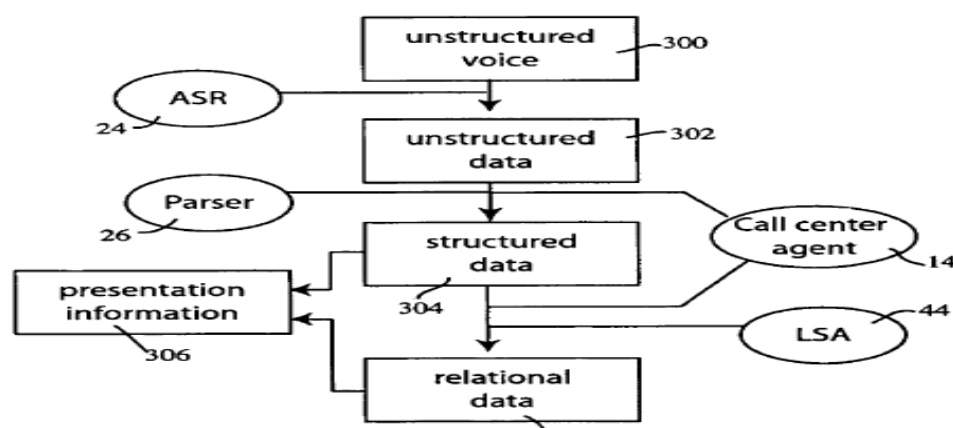


Fig2: The above figure explains the basic process of unstructured voice information processing.

Source: Assistive call center interface, Applebaum et al, September, 2006

The output of these parsers is a structured data that is now transferred to Data Manager which employs an associated query generator that is connected to various databases like product database, prior call history database, etc. This is done to get any additional information from the unstructured voice input of the caller or to compare the information provided by the caller with the prior call history. Such information is also used by data manager to make queries

asking additional information from the caller. Data manager is also associated with a presentation logic that is used to generate various reports and auto displays showing product or service information related to keywords found in the caller's speech. The reports can be in the form of a running transcript of the recognized speech from the incoming caller highlighted with keywords.

Apart from its use in call processing information in call centers, this invention also provides a robust system of transferring expertise and knowledge between call center agents by storing these database queries along with the transcript of the calls. This can be referred by the same or some other agent if a similar query is raised.

2.1.3 The Process of Guy Alon

[3] Guy Alon (2005) described another method which is context independent and can analyze a continuous and spontaneous conversation between two or more people that includes a lot of vocabulary. This method is called Key Word Spotting (KWS). Key Word Spotting is an application of Automatic Speech Recognition (ASR), KWS uses ASR engine to analyze a large number of calls and conversations between two or more people and uses unlimited vocabulary for this purpose. The information obtained from KWS is valuable as it helps in analyzing the caller's behavior and state of mind. Guy Alon has described mainly three approaches to KWS in speech analytics. These are as follows:

- LVCSR based KWS: It's a two stage process where the speech is transcribed into text in the first stage using Large Vocabulary Continuous Speech Recognition followed by key word search in the second phase.
- Phoneme Recognition based KWS: This is also a two staged process. In the first stage, speech is transformed into a sequence of phoneme, followed by a search for phonetically transcribed key words. This approach works for a lower linguistic level which is called the phoneme recognition. Phoneme are the basic speech units that when combined can represent any word in the language.
- Word Recognition based KWS: It's a one stage process where KWS based engine looks for keywords in the speech stream based on a target phoneme sequence representing the keyword.

[4] Call centers also use Computer Telephony Integration (CTI) system in order to perform compliance check i.e. to check whether the agent has made some predetermined statements related to a call transaction. It is also used for security mechanism via the process of speaker authentication. CTI extracts audio signal from the telephone calls and perform real time voice analytics.

2.2 Use of R in Voice Analytics

[5] Like many other technologies, R programming also finds its use in voice analytics. R packages like tuneR and audio make it possible to record, import, modify and export audio objects.

tuneR and seewave packages can be used to analyze sound. Many complimentary functions are offered by these packages and they can be used to extract and compare relevant amplitude, temporal, phase and frequency parameters. The package seewave is now widely used in bioacoustics, a discipline of life sciences that analyze animal sounds. Some new packages like playitbyr and audiolyzR can be used for sonification. Sonification can be compared to visualization, as data can be mapped to graphics using visualization, similarly data can be transferred into non speech sound using the process of sonification. Playitbyr is used to map data onto sonic parameters like pitch, tempo and rhythm.

2.3 Voice Analytics Application in various fields:

Voice Analytics Market is growing at a very good pace and its importance is realized in many fields. The first professional application of speech technology was in medical field. In many studies, the accuracy of voice recognition technology has been claimed as high as 99%. Also, many organizations nowadays use cloud based and real time analytics solutions. Because of this, voice analytics finds its application across many verticals where it is used to fulfill many unique purposes. Some of the verticals and the use of voice analytics in these verticals are briefly described as follows:

2.3.1 In Call centers: As mentioned previously, voice analytics hold major importance for this market and hence is extensively used for various types of analysis that are helpful for the business. Call centers use different types of techniques; some of them are described above, to perform voice analysis

[6] Another technique that is used in Call centers for voice analytics is “Multi Pass Analytics” which is basically a two tier process wherein the first tier recognition server receives the audio component and converts it into a text format using phonetic analysis or LVCSR analysis. The text format is then sent to the second tier recognition server which is basically an analytic component and can be used to analyze the quality of communication between the agent and the caller. In other cases this technology can also be used for speech recognition. In some embodiments, first tier can also perform precursory text analytics after converting audio signal to text format. Upon recognition of the desired search term, first tier recognition server sends only a part of the communication to second tier recognition server for thorough analysis. This embodiment makes the overall process fast. This technology also helps in speaker identification, speaker verification and determination of emotions of speakers that makes it further attractive for call centers.

[7] Phonetic Analysis is one more technology that is frequently used in call centers and has various applications. Phonetic analysis is the systematic study of the sound of human speech. There are approximately 400 sounds or “phonemes” within all human languages and by collecting them, organizations are able to capture a true record of what is said in an audio track that can be searched more accurately and flexibly than human analysts could otherwise do on their own. The various applications of phonetic analysis in call centers are:

- 1) **Recognition:** Voice command recognition, substitutes for button type automated response, enables human-computer interaction. Some example of solution providers are Bridgetec (Catch You), Nuance Recognizer, etc.
- 2) **Analytics:** Helps transform unstructured phonetic data into well structure text-based data through the Speech To Text engine for phonetic data of customers inbound to call centers; analyzes and categorizes customer interaction data collected in call centers; provides a cross-channel analytics platform enabling companies to transform valuable but hidden information in customer interaction into business results. Examples of solution providers are Verint Systems (Impact 360), NICE Systems (Interaction Analytics), etc.
- 3) **Authentication:** Uses the unique voice tones of calling customers; call center agents to verify identities; transparently retrieves the biometric voice characteristics required for verification within seconds regardless of what is said, accent, or language; verifies customer identity by comparing calling customer voice tones with those recorded in a database. Some examples of solution providers are Bridgetec (Catch Who), Nuance (FreeSpeech), etc.

This technique provides a lot of advantages to the call centers. Some of them are as follows:

- 1) Improved customer satisfaction, customer accessibility, improved customer experience, reduced customer churns; improved operating efficiency through increased self service; reduced communication fees.
- 2) Reduced talk time and improved first-call resolution; reduced agent overload and supervision.
- 3) Identify customer related issues and obtain internal customer satisfaction data; help gain useful feedback from current markets and trends; implement marketing strategy planning.
- 4) Reduced call center fraud.

2.3.2 In Military: [8], [9] Various speech and language technology components like speech coders, voice controlled C2 systems, speaker and language recognition, translation systems and automatic training suites are widely used in military communications, for command and control, in intelligence as well as in various training systems. Even US military has added speech recognition technology into its Leeds lifecycle engineering system. This system holds data on design, parts inventories, operational histories and maintenance instructions. This speech interface gives engineers a hand free access to data when they perform maintenance tasks and also helps in recording inspections through voice input.

2.3.3 In Aviation: In aviation, speech recognition products are mainly used for GPS that enables pilots to enter GPS waypoints 10 times faster than using control knobs, which helps them to keep their attention on the sky. These devices are mainly used by pilots flying Beechcraft Baron, Piper Arrow, and high-performance FJ-4 Fury. The technology developed by VoiceFlight has become the first Federal Aviation Administration (FAA)-certified speech recognition product for use in civilian aircraft.

One more tested system for civil aircraft is Honeywell Innovative Prototyping Environment (HIPE). Using this system, the pilots are able to operate a traditional multipurpose control display unit from a tablet interface to incorporate new concepts. Honeywell is using HIPE to access its various possibilities in terms of usability, efficiency and safety in real airborne scenarios. For this it is testing the flight voice recognition system on an Embraer ERJ170 aircraft.

2.3.4 In Financial Market: [10] One of the applications of voice analytics in financial market is identifying intentional deception in financial reporting. Many sophisticated market participants like institutional investors and analysts have been unsuccessful in detecting these financial frauds making the situation more serious and dangerous. Hence it became increasingly important to develop a framework that can timely report such financial frauds and could help investors, auditors, analysts and regulators. Many psychologists in their research have claimed that there are many verbal linguistic cues (e.g., speech content) and non verbal cues (e.g., tone of voice, facial expression or gestures) that can detect deception. One such vocal marker of deception (vocal marker is a type of non-verbal cue) is Cognitive dissonance which is a state of psychological arousal and discomfort occurring when an individual takes actions that contrast with a belief, such as cheating while believing one to be honest. (Source: Hobson J.L. et al [2012]. Analyzing Speech to detect Financial Misreporting)

Researchers collected speech samples of CEOs during their interaction with analysts and investors over a conference call and tried to find out vocal cognitive dissonance markers in their speech using automated vocal emotional analysis software based on Layered Voice Analysis (LVA) technology. LVA is an emerging technology used in the financial markets to detect fraud. With this technology researchers concluded that there is a direct link between the presence of vocal dissonance markers in CEO's speech and their deception towards financial misreporting.

2.3.5 In Healthcare: [11] Speech analytics has found its application in healthcare as well. It can be used to detect Parkinson's disease in a patient. Parkinson's disease is a disorder of Central Nervous System and in 90% of the cases, people with Parkinson's disease has speech or voice disorders. By analyzing voice samples of healthy people and those who are suspected to have Parkinson's disease, it can be confirmed whether a person suffers this disease or not. It is also used to detect depression and other mental issues. One such effort has been made by a Massachusetts based company, Cogito Corporation. With the help Cogito's Social Signal Processing (SSP)

technology platform, the company has developed inexpensive, brief and scalable speech behavior analytics tools that can be accessed by internet, IVR or live-to-screen for depressed voices. This technology is also able to access cues in an individual's natural speech to provide accurate, non –invasive monitoring to support intervention for psychological issues.

2.4 Growth of Voice Analytics market:

The increase in the number, the importance of voice in multichannel world, the increased focus on compliance, risk management and competitive market intelligence; all of these reasons has led to an upsurge of the voice analytics market. Also, it gives a competitive advantage to the organizations that uses this technology and has numerous applications in various fields as described in the previous section of this paper. A number of vendors has now realized the business advantages of this technology and providing various speech related products to different organizations. [12] Some of the major vendors are listed in the table as below:

Vendor	Product	Approach Used
NICE	Nice Interaction Analytics	Phonetic Indexing and Transcription
VERINT	Impact 360 Speech Analytics	Audio Indexing
HP Autonomy	HP IDOL(Intelligent data Operating Layer)	Meaning Based Computing (MBC). MBC stresses relevance along with accuracy
Nexidia	Neural Phonetic Speech Analytics	Automatic Speech Recognition (ASR), phonetic indexing and word-level transcription
Genesys (UTOPY)	Speech and Text Analytics (formerly, SpeechMiner)	Direct Phase detection and Transcription

Table 1: Source: Deciphering Voice of Customers through Speech Analytics, 2015, Computaris

Although the number of organizations using speech solution is still less, but DMG consulting projected the growth of this market to be around 18% in 2015 and 16% in 2016 and 2017.

[13] According to a report published by DMG Consulting, Nice continued to be the leader in providing speech analytics solutions in terms of market share and its share has gone up from 30% in 2013 to 34.5% in 2014. Verint

6th International Conference on Science, Technology and Management

India International Centre, New Delhi

04 December 2016 , www.conferenceworld.in

(ICSTM-16)

ISBN: 978-93-86171-16-0

occupied the second position followed by HP. The table below shows the market share and activity of 16 best speech analytics vendors in the world as of March 31st, 2015:

Figure: Speech Analytics Activity as of March 31 st , 2015			
	Total		Seats
Vendor	Customers	Seats	Market Share
Nice	704	1,210,000	34.5%
Verint	624	834,732	23.8%
HP	254	400,000	11.4%
CallMiner	174	321,657	9.2%
Nexidia	195	305,000	8.7%
Interactive Intelligence	138	70,934	2.0%
Mattersight	56	56,000	1.6%
Avaya (Aurix)	67	53,372	1.5%
Genesys (UTOPY)	55	51,000	1.5%
Calabrio	63	23,485	0.7%
inContact	85	17,087	0.5%
Aspect	27	16,500	0.5%
Zoom International	45	8,000	0.2%
Castel	28	7,617	0.2%
CTI Group	24	7,000	0.2%
OnviSource	31	3,994	0.1%
Other	129	169,319	4.8%
Subtotal	2,699	3,55697	N/A
Less: Double Count	(67)	(47,902)	(1.4%)
Total	2,632	3,507,795	100.00%

Table 2: Source: DMG Consulting LLC, June 2015

[14], [15] In one of the recent reports by SYS.CON Media and The Movers News speech analytics market is expected to grow from USD 589.1 million in 2015 to USD 1.60billion by 2020, at an estimated Compound Annual

6th International Conference on Science, Technology and Management

India International Centre, New Delhi

04 December 2016 , www.conferenceworld.in

(ICSTM-16)

ISBN: 978-93-86171-16-0

Growth Rate (CAGR) of 22% from 2015 to 2020. The professional service center is expected to capture a major portion of this growing speech analytics market.

North America captures the largest market and Latin America is the fastest growing market in terms of CAGR. Apart from obtaining a competitive advantage, the latest trend of cloud based speech analytics solution and real time speech analytics solution has given a further boost to the demand of speech analytics based solutions across many verticals. The market is even proliferating rapidly in Asia Pacific region due to increasing economic growth and cloud adoption in the region.

On seeing the potential of this market, many investors are showing interest and funding start-ups and other vendors that provide speech related solution. For example, recently CallMiner came up with a new product called as Eureka Live that can help call center workers analyze customer speech in real time, track their own performance and guide them to perform efficiently. This was all possible because CallMiner technologies have attracted many investors like NewSpring Capital that has invested \$12 million in the company alongside current investors making the total investment money to around \$ 16 million.

Even agencies like the Central Intelligence Agency (U.S.A) are funding entrepreneurs that have ideas that might help track terrorists and dangerous foreign spies. In 1999, the CIA created In-Q-Tel to fund promising technologies start-ups that might prove beneficial in defending the country, because by this time they realized that commercial technologies were moving far ahead compared to government agencies in providing innovative solutions in fields like communication and biotechnology.

All of these points prove that how important this new technology has become which has lead to the fast pace growth of voice analytics market in a very short span of time.

III CONCLUSION

The paper reviews the progress made in the field of voice analytics over the years and summarizes the technique, applications and the market growth of voice analytics. The paper does not include any analysis of sample voice data as it is difficult and expensive to get licensed software that can be used for this purpose. The purpose behind writing this paper is to draw attention of not just the technical experts but also the entrepreneurs so that both can work in sync to develop this technology that can be beneficial in many ways. Many benefits of this technology in various fields like healthcare, call center, military, financial markets, etc. is described in this paper and this shows that only in one decade voice analytics has attained such critical positions in various fields. Voice analytics can provide a kind of business intelligence that can help organizations to learn more about their customers, their products and services, etc. which can help organizations to develop products and services that can improve customer satisfaction. Organizations are currently using various analytical methods like text analytics and sentiment analytics that take input data from various social media in order to understand about their customer. Voice analytics is one step ahead

6th International Conference on Science, Technology and Management

India International Centre, New Delhi

04 December 2016 , www.conferenceworld.in

(ICSTM-16)

ISBN: 978-93-86171-16-0

of this and is better able to produce a complete picture of their business environment. Not only this, its potential to identify fraud and detect terrorism can provide a solution to deal with threats and security. Hence, it is imperative to draw attention to voice analytics and integrate it with various other fields to develop superior technologies.

REFERENCES:

- [1] Chan, M., Donohue, P., & Uckun, S. (2003). *U.S. Patent No. 6,600,821*. Washington, DC: U.S. Patent and Trademark Office.
- [2] Applebaum, T., & Junqua, J. C. (2006). *U.S. Patent No. 7,103,553*. Washington, DC: U.S. Patent and Trademark Office.
- [3] Alon, G. (2005). Key-word spotting—the base technology for speech analytics. *Natural Speech Communications*.
- [4] Pickering, R., Moussalli, J., & Unitt, A. (2013). *U.S. Patent No. 8,600,756*. Washington, DC: U.S. Patent and Trademark Office.
- [5] Sueur, J. R as a Sound System. *Muséum national d'Histoire naturelle*
- [6] Watson, J. (2010). *U.S. Patent No. 7,752,043*. Washington, DC: U.S. Patent and Trademark Office.
- [7] Shim, J. P., Koh, J., Fister, S., & Seo, H. Y. (2016). Phonetic analytics technology and big data: real-world cases. *Communications of the ACM*, 59(2), 84-90.
- [8] (2006, April 18). US military to add voice interface to engineer's mobile IT. TechTarget. Retrieved from <http://www.computerweekly.com/news/2240062839/US-military-to-add-voice-interface-to-engineers-mobile-IT>
- [9] Masterson, M. (2012, June 5). Military, Insurers Use Speech Solution to Detect Depression. Speech Technology. Retrieved from <http://www.speechtechblog.com/2012/06/05/military-insurers-use-speech-solution-to-detect-depression#axzz476Ve0ejk>
- [10] Hobson, J. L., Mayew, W. J., & Venkatachalam, M. (2012). Analyzing speech to detect financial misreporting. *Journal of Accounting Research*, 50(2), 349-392.
- [11] Shrivan R. A., Tahami, E.(2011).Voice analysis for detecting Parkinson's disease using genetic algorithm and KNN classification method. *Biomedical Engineering (ICBME), 2011 18th Iranian Conference of*, Tehran, 278-283
- [12] Computaris. (2015).*Deciphering voice of Customer through Speech Analytics*.
- [13] DMG Consulting LLC (2015). *Speech and text analytics product and market report reprint*. New Jersey: DMG Consulting.
- [14] Speech Analytics Market- Global Forecast to 2020 – Increase in customer interaction data through various channels- Research and Markets. (2016, April 20). *SYS.CON MEDIA*
- [15] Speech Analytics market to see a growth of 22% CAGR to 2020 driven by growing adoption of contact centers. (2016, April 22). *The Movers News*