Research Journal of Applied Sciences, Engineering and Technology 9(10): 862-868, 2015
DOI:10.19026/rjaset.9.2636
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: November 19, 2014 Accepted: January 8, 2015 Published: April 05, 2015

Research Article

A Study to Increase the Quality of Financial and Operational Performances of Call Centers using Speech Technology

[1]R. Manoharan, [2]R. Ganesan and [3]D. Venkata Subramanian
[1]Research and Development Center, Bharathiar University, Coimbatore-641 046, India
[2]Sri Venkateswara College of Computer Applications and Management, Coimbatore-105, India
[3]Saveetha School of Engineering, Thandalam, Chennai-602 105, India

Abstract: Everyone knows technology and automation are not the solutions to every business problem. But when used for the right reasons-and deployed and maintained wisely-speech based contact center applications can be good for the customers as well as business, if the money and time spent to implement and maintain contact center business. Speech based application is an experimental conversational speech system. Experience with redesigning the system based on user feedback indicates the importance of adhering to conversational conventions when designing speech interfaces, particularly in the face of speech recognition errors. Study results also suggest that speech-only interfaces should be designed from scratch rather than directly translated from their graphical counterparts. This paper examines a set of challenging issues facing speech interface designers and describes approaches to address some of these challenges. Let us highlight some of the specific constrains involved in this process of using speech technology in the main stream of business in general and of a Call Center specific and being resolved through the Industrial process. So the real challenge is "Developing a new Business Process Model for an Industry application, for a Call Center specific. And that paves the way to design and analyze the "Financial and Operational performance of call centers through the business process model using speech technology".

Keywords: Call center operations, cost reduction, QoS, SEL, speech user application

INTRODUCTION

Speech and natural language understanding are the key technologies that will have the most impact in the next 15 years (Bill, 2004). Speech recognition is the most compelling form of self-service for companies large and small because it's satisfying for customers and cost effective for enterprises. A well designed speech application gives customers the information they want, when they want it. It doesn't require customers to wait for their PC to "boot" or to navigate a Web site in the hope of finding answers. It doesn't force customers to remember and enter numbers into their touch-tone phone to obtain only a portion of the required information. Nor does it force callers to repeatedly hang up and start over, a common occurrence in many Interactive Voice Response (IVR) environments. A well-tuned speech recognition application anticipates questions and provides accurate answers in a user-friendly manner that promotes a positive customer experience. The economic slowdown has accelerated the need to automate a larger percentage of service requests. With few exceptions, companies can no longer afford to provide live assistance to all callers; it's simply too expensive. According to Gartner Group, a typical service call costs $5.50, as compared to $0.45 for a call handled by an IVR. Companies that use speech recognition will improve service quality and enhance the customer experience while reducing costs, giving them a distinct competitive advantage (Fig. 1).

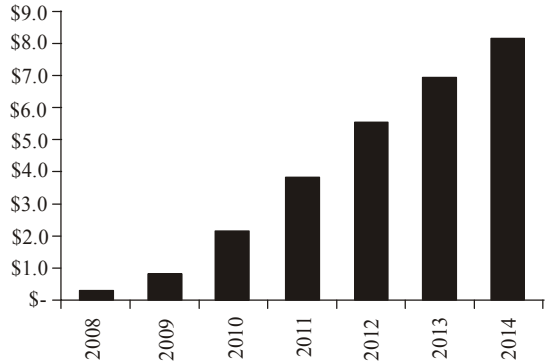

Fig. 1: Mobile applications revenue (Zach Epstein, 2011)

According to IHS-owned market research firm iSuppli, revenues from major mobile app stores will grow 77.7% to $3.8 billion this year.

Every day, tens of millions of help-desk calls are recorded at call centers around the world. As part of a typical call center operation a random sample of these calls is normally re-played to human monitors who

Corresponding Author: R. Manoharan, Research and Development Center, Bharathiar University, Coimbatore-641 046, India

score the calls with respect to a variety of quality related questions, e.g.:

- Was the account successfully identified by the agent?
- Did the agent request error codes/messages to help determine the problem?
- Was the problem resolved?
- Did the agent maintain appropriate tone, pitch, volume and pace?

This process suffers from a number of important problems: first, the monitoring at least doubles the cost of each call (first an operator is paid to take it, then a monitor to evaluate it). This causes the second problem, which is that therefore only a very small sample of calls, e.g., a fraction of a percent, is typically evaluated. The third problem arises from the fact that most calls are ordinary and uninteresting; with random sampling, the human monitors spend most of their time listening to uninteresting calls.

Speech recognition is used to transcribe 100% of the calls coming in to a call center and default quality scores are assigned based on features such as key-words, key-phrases, the number and type of hesitations and the average silence durations. The default score is used to rank the calls from worst-to-best and this sorted list is made available to the human evaluators, who can thus spend their time listening only to calls for which there is some a-priori reason to expect that there is something interesting.

The automatic quality-monitoring problem is interesting in part because of the variability in how hard it is to answer the questions. Some questions, for example, "Did the agent use courteous words and phrases?" are relatively straightforward to answer by looking for key words and phrases. Others, however, require essentially human-level knowledge to answer; for example one company's monitors are asked to answer the question "Did the agent take ownership of the problem?" It is suggested that there is a set of 31 questions that are used to evaluate call-quality. Because of the high degree of variability found in these calls, we have investigated two approaches:

- Use a partial score based only on the subset of questions that can be reliably answered.
- Use a maximum entropy classifier to map directly from ASR-generated features to the probability that a call is bad (defined as belonging to the bottom 20% of calls).

We have found that both approaches are workable and we present final results based on an interpolation between the two scores. These results indicate that for a fixed amount of listening effort, the number of bad calls that are identified approximately triples with our call-ranking approach. Surprisingly, while there has been significant previous scholarly research in automated call-routing and classification in the call center, e.g., (1, 2, 3, 4, 5), there has been much less in automated quality monitoring per sec.

Speech recognition benefits in a call center-operational: Speech recognition has traditionally been viewed as a contact center productivity tool to increase the quality of service organizations because that is where the financial payback has been the greatest. But the technology has already proven effective in improving productivity and generating revenue outside of contact centers, when used throughout the enterprise. On the productivity side, speech recognition yields the following benefits:

- Reduces calls to live agents
- Shortens call lengths
- Reduces call hold times
- Decreases call abandonment rates

Speech recognition justification criteria in a call center-financial: Technology investments must be justified based on hard and quantifiable benefits, as these are the only measures acceptable to Chief Financial Officers (CFOs) today. However, the technology selection decision should reflect both hard and soft benefits, as both are relevant to the department's performance. Cost centers can justify speech investments on: The five dominants Production, Productivity, Waste Control, Cost Control and Cost Reduction, are improving reduction in agents, supervisors, trainers, QA specialists.

Cost reduction: Reducing the number of calls, agent talk time, line charges (from reduced agent talk time and hold time), hiring and training costs.

Cost avoidance: Eliminating the need to purchase additional hardware or software to handle incremental calls.

Profit centers can also include incremental revenue in their Return of Investments (ROI) analyses, even though CFOs will not accept soft benefits, as they can't be easily quantified or attributed to a particular investment. Speech recognition soft benefits that an enterprise should still evaluate during the selection process include:

- Reduction in the number of abandoned calls
- Reduction in customer call backs
- Reduction in call center hardware and software
- Reduction in agent attrition
- Increase in customer satisfaction

This study describes an automated quality-monitoring system that addresses these problems

through a speech user application business model proposed for a call center.

LITERATURE REVIEW

Most companies understand the benefits of the call center office, where business gets done far beyond the confines of brick-and-mortar buildings. But even with this understanding, most still remain in bondage to the keyboard. For too long, the computer keyboard has been the primary means for turning ideas into workflow. The missing ingredient is voice technology; when you add speech processing to the mix, the call centers finally reaches its full potential.

Businesses around the globe are discovering that voice technology dramatically boosts efficiency and productivity in ways that text-heavy solutions cannot. By greatly improving response time and workflow, speech processing has become a significant competitive advantage-and that speaks volumes to corporate leaders worldwide (Fugate, 2013).

Perhaps one of the best advantages of voice technology is that its use is not limited to only a select few professions. Speech processing already is increasing efficiency in many fields of business. Here are a few real-life examples (www.iq-services.com).

Building relationships, fostering teamwork: In addition to boosting productivity, voice technology is also a natural for building business relationships. In fact, voice processing nurtures relationships far better than typing ever will. Suppose you were to visit a client and discover that his son is graduating from Yale in six weeks. You could capture that information by quickly sending a voice file to your support staff. Long after the face-to-face meeting, the client is pleasantly surprised to get your graduation card. It is an efficient way to build strong business connections.

By adding voice technology capabilities, the call center office has finally come of age. It frees companies from computers and keyboards, allowing work to be done anywhere, at any time. With speech processing, both the business person and the support team are seamlessly united to achieve remarkable efficiencies.

According to a report from DMG Consulting, “an astounding 28.1% of companies not currently using Interactive Voice Response (IVR) systems are looking for a voice self-service solution to help meet their goals.” (www.iq-services.com). The report goes on to say that “the trend toward increased adoption of IVR solutions is expected to continue even after the economy recovers.” That is why using a trusted partner with worldwide testing capacity to help assure the applications deliver the highest quality of service possible is so important. In this latest Best Practices Series on contact center business applications that use speech, read how IQ Services can help test those speech based contact center business applications to ensure they are done right, which will help protect your investment for the short and long terms. (www.iq-services.com).

Building a sound social presence (www.speechtechmag.com/Articles/Editorial/Feature/): Many people scratched their heads when Twitter launched in mid-2006, scoffing at the idea of condensing any meaningful information into 140-character tweets. Soon, they found that a lot could be said in a little space and the social networking site quickly became one of the most popular properties on the Web.

Now another fast-emerging social media trend has people scratching their heads: Voice-based micro blogging, which enables users to record short voice messages and share them with their social media followers, is rising from the ashes after all but completely burning out a few years ago.

Some have approached the micro blogging trend with skepticism, while others are celebrating it as the next great business tool for communicating with customers, partners and employees and for creating highly engaged social communities.

"It can be a good tool to make (customer) interactions more personal," says Kimberly Chau, an AMI-Partners marketing associate focused on social media. "It can give a company a personality and it's very easy to do."

"From a personal standpoint, anything that creates more trust online is good and I could see this building trust," says Michael Fauscette, senior analyst and head of the Software Business Solutions Group at IDC. "From a business standpoint, this is perhaps a medium to build more customer engagement. It can also be seen a marketing angle."

"Engaging with potential and existing customers or clients via the spoken word can be much more effective and effectual" than text-only posts, adds Rob Proctor, CEO of Audio boo, a London-based voice messaging and micro blogging platforms provider.

"Audio does a lot more than text-only," says Taylor Bollmann, CEO of San Francisco-based Yiip, a voice messaging platform provider. "If a product is something the consumer is really into, getting emotionally rich and entertaining content (in the form of a voice file) is really cool."

Media and entertainment companies have already started. The BBC, Sky News and the Royal Opera House are frequent users of Audioboo and they encourage listeners to respond to the online audio with their own voice comments and share those comments with their friends and family, creating ongoing dialogues and energizing their fan bases.

Analytics and compliance increase call recording investments (www.speechtechmag.com/Articles/

Editorial/FYI/): Global sales of contact center interaction recording systems are poised to grow from $813 million in 2011 to $1.2 billion in 2018, according to a new report from Pelorus Associates.

Dick Bucci, author of the report and principal analyst at Pelorus, defines recording systems as those that include speech analytics, e-learning, data analytics, quality management and other products associated with the core recording system.

Some of the growth in the recording systems market can be tied to companies that have old infrastructures and are interested in adding new capabilities, such as speech and data analytics. Another driver for growth is the issue of compliance, Bucci says.

DMG Consulting defines voice recording systems a bit differently. According to Donna Fluss, founder and president of DMG Consulting, recording is only one aspect of much larger workforce optimization suites. The other modules are quality assurance, coaching, e-learning, performance management, surveying, speech analytics, workforce management, desktop analytics and text analytics.

"Projections for recording are actually pretty small, if we're looking at just recording," says Fluss, who estimates growth at 3% in 2012, 3% in 2013, 4% in 2014 and 2% in 2015.

Fluss also points out that recording systems for both contact centers and non-contact centers are not just being sold by the Workforce Optimization (WFO) vendors, but almost every single cloud-based contact center infrastructure vendor. There are also standalone recording vendors, WFO vendors and now hosted vendors that are selling recording as part of their core capabilities.

NICE Systems and Verint Systems are major players in this arena and according to Fluss, account for 65 to 90% of all the different slices in the WFO market. Based on the numbers Fluss has seen for the end of fiscal year 2012, it appears that Avaya is now going to be the third largest WFO vendor in the market. Vendors are increasingly putting their energy into the applications that use recordings to do other things like speech analytics.

MATERIALS AND METHODS

With this background of study it is focused to evaluate financial and operational performance of call centers by increasing quality of services, in two stages. Those are:

- All input and output transactions are captured and recorded as error free content for a domain specific application for a call center through a Business Process Model using an algorithm named Speech Empowerment Lemma (SEL) (Manoharan *et al*., 2009) (i.e., The algorithm made of tools like Fast Fourier Transform and validated through Power Law). The domain specific captured sample content then be used by software tools like Microsoft Speech Server (MSS) and front-end language Salt Application Language Tags (SALT).
- And a survey used to collect feedback from that call center operators and customers through a well composite comprehensively designed questionnaire form and to exhibit how the operational performance and ultimately the financial performance compared with the legacy systems. For which a multivariate-Regression Analysis is done using tools like Statistical Package for Social Service (SPSS).

ASR for call center transcription:

Data: The speech recognition systems were trained on approximately 300 h of 6 kHz, mono audio data collected at one of the call centers located in Coimbatore, India. The audio was manually transcribed and speaker turns were explicitly marked in the word transcriptions but not the corresponding times. In order to detect speaker changes in the training data, it is a forced-alignment of the data and chopped it at speaker boundaries. The test set consists of 50 calls with 113 speakers totaling about 3 h of speech.

Speaker independent system: The raw acoustic features used for segmentation and recognition are Perceptual Linear Prediction (PLP) features. For the speaker independent system, the features are mean-normalized on a per speaker basis.

Every 9 consecutive 13-dimensional PLP frames are concatenated and projected down to 40 dimensions using LDA+MLLT. The SI acoustic model consists of 50 K Gaussian strained with MPE and uses a quinph one cross-word acoustic context. The techniques are the same as those described in (Soltau *et al*., 2005).

Incremental speaker adaptation: In the context of speaker-adaptive training, two forms of feature-space normalization are used: Vocal Tract Length Normalization (VTLN) and feature-space MLLR (fMLLR, also known as constrained MLLR) to produce canonical acoustic models in which some of the non-linguistic sources of speech variability have been reduced. To this canonical feature space, it is then applied a discriminatively trained transform called Fmpe (Povey *et al*., 2005). The speaker adapted recognition model is trained in this resulting feature space using MPE (Povey *et al*., 2005).

It is distinguished between two forms of adaptation: Off-line and incremental adaptation. For the former, the transformations are computed per conversation-side using the full output of a speaker independent system. For the latter, the transformations are updated incrementally using the decoded output of the speaker adapted system up to the current time. The

Table 1: ASR results depending on segmentation/clustering and adaptation type

Segmentation/clustering	Adaptation	WER (%)
Manual	Off-line	30.2
Manual	Incremental	31.3
Manual	No adaptation	35.9
Automatic	Off-line	33.0
Automatic	Incremental	35.1

Table 2: Accuracy for the question answering system

Accuracy	Top 20%	Bottom 20%
Random	20%	20%
QA	41%	30%

speaker adaptive transforms are then applied to the future sentences. The advantage of incremental adaptation is that it only requires a single decoding pass (as opposed to two passes for off-line adaptation) resulting in a decoding process which is twice as fast. In Table 1, we compare the performance of the two approaches. Most of the gain of full offline adaptation is retained in the incremental version (Povey *et al*., 2005).

Call ranking: This section presents automated techniques for evaluating call quality. These techniques were developed using a training/development set of 676 calls with associated manually generated quality evaluations. The test set consists of 195 calls. The quality of the service provided by the help-desk representatives is commonly assessed by having human monitors listen to a random sample of the calls and then fill in evaluation forms. The form for contact center Help Desk contains 31 questions. A subset of the questions can be answered easily using automatic methods, among those the ones that check that the agent followed the guidelines e.g.:

- Did the agent follow the appropriate closing script?
- Did the agent identify herself to the customer?

But some of the questions require human-level knowledge of the world to answer, e.g.:

- Did the agent ask pertinent questions to gain clarity of the problem?
- Were all available resources used to solve the problem?

These were able to answer 21 out of the 31 questions using pattern matching techniques. For example, if the question is "Did the agent follow the appropriate closing script?", it is searched for "THANK YOU FOR CALLING", "ANYTHING ELSE" and "SERVICE REQUEST". Any of these is a good partial match for the full script, "Thank you for calling, is there anything else I can help you with before closing this service request?" Based on the answer to each of the 21 questions, it is computed a score for each call and use it to rank them. A call is labeled in the test set as being bad/good if it has been placed in the bottom/top 20% by human evaluators. The accuracy of our scoring system on the test set is reported by computing the number of bad calls that occur in the bottom 20% of our sorted list and the number of good calls found in the top 20% of our list. The accuracy numbers can be found in Table 2.

Maximum ranking: Another alternative for scoring calls is to find arbitrary features in the speech recognition output that correlate with the outcome of a call being in the bottom 20% or not.

The goal is to estimate the probability of a call being bad based on features extracted from the automatic transcription. To achieve this it is built a maximum entropy based system which is trained on a set of calls with associated transcriptions and manual evaluations. The following equation is used to determine the score of a call_using a set of predefined features:

$$P(class/C) = \frac{1}{Z} exp(\sum_{i=1}^{N} \lambda_i f_i(class, C)) \quad (1)$$

where, *class* € {bad, not-bad}, Z is a normalizing factor, $f_i()$ are indicator functions and $\{\lambda\}_{\{i\text{-}1,N\}}$ are the parameters of the model estimated via iterative scaling (Fugate, 2013).

End-to-end system performance: This section describes the user interface of the automated quality monitoring application. As explained aibove, the evaluator scores calls with respect to a set of quality-related questions after listening to the calls. To aid this process, the user interface provides an efficient mechanism for the human evaluator to select calls, e.g.:

- All calls from a specific agent sorted by score
- The top 20% or the bottom 20% of the calls from a specific agent ranked by score
- The top 20% or the bottom 20% of all calls from all agents

The automated quality monitoring user interface is a MICROSOFT web application that is supported by back-end databases and content management systems (Chu-Carroll and Carpenter, 1999). The displayed list of calls provides a link to the audio, the automatically filled evaluation form, the overall score for this call, the agent's name, server location, call id, date and duration of the call (Fig. 2). This interface now gives the agent the ability to listen to interesting calls and update the answers in the evaluation form if necessary (audio and evaluation form illustrated in Fig. 3). In addition, this interface provides the evaluator with the ability to view summary statistics (average score) and additional information about the quality of the calls.

Human vs. computer rankings: As a final measure of performance, (Fig. 4).

It is presented a scatterplot comparing human to computer rankings. No calls that are scored by two

Fig. 2: Display of selected calls (Berger *et al*., 1996)

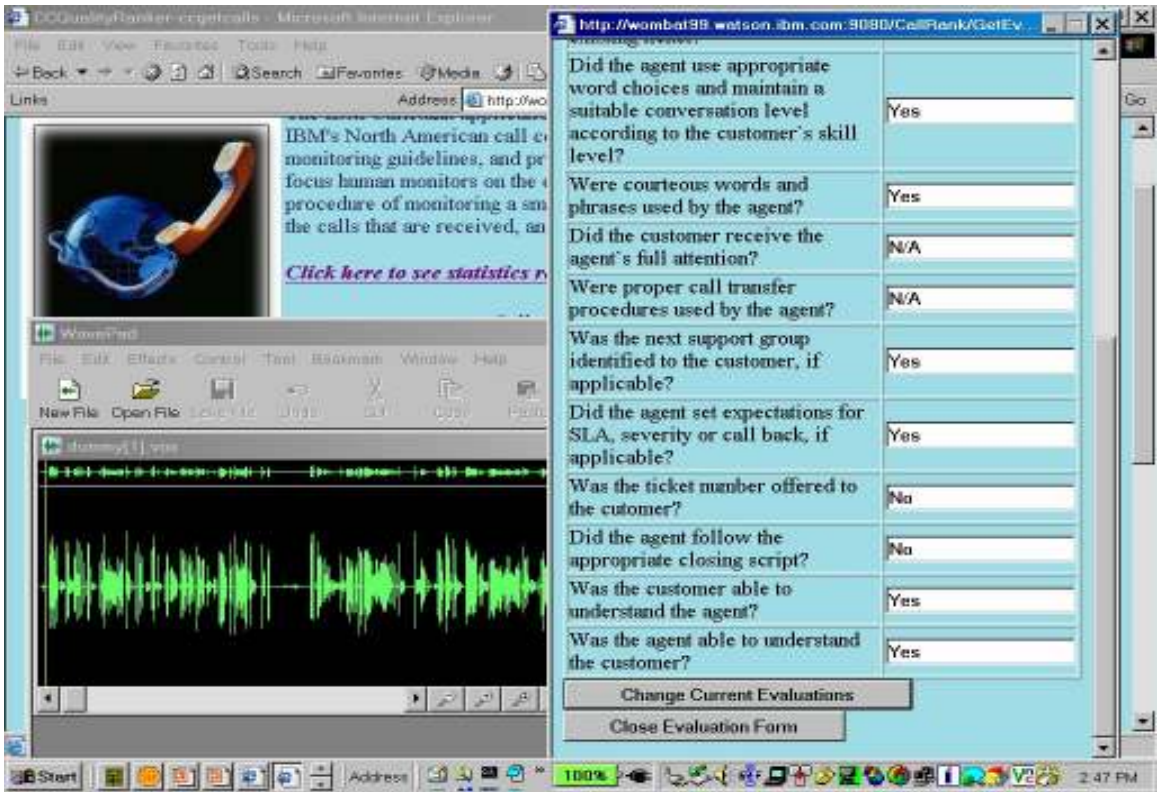

Fig. 3: Interface to listen to audio and update the evaluation form (Berger *et al*., 1996)

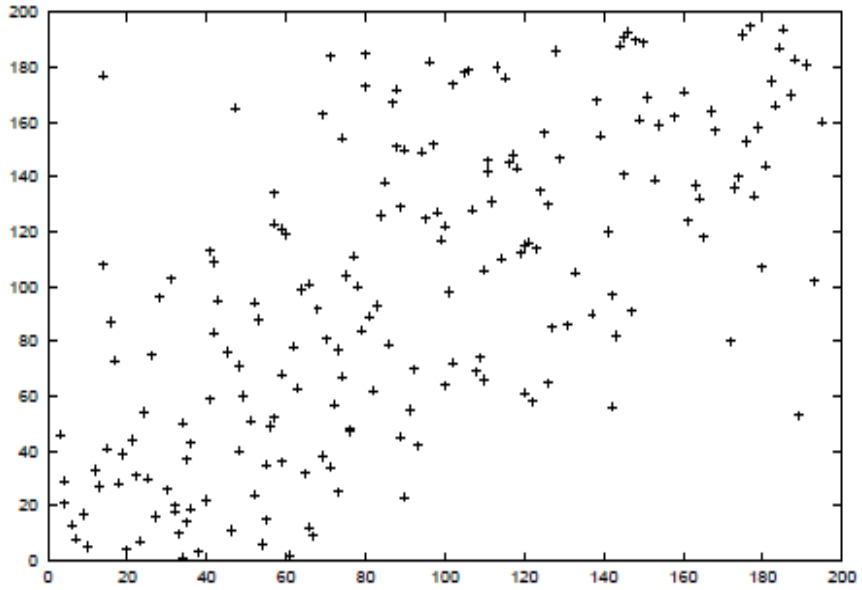

Fig. 4: Scatter plot of human vs. computer rank (Berger *et al*., 1996)

humans, so it is not presented a human-human scatterplot for comparison.

RESULTS AND DISCUSSION

The system can either be used to replace human monitors, or to make them more efficient. The results show that it can triple the efficiency of human monitors in the sense of identifying three times as many bad calls for the same amount of listening effort.

Speech technology is being used in a call center to increase revenue by automating activities that previously required agent assistance, such as a credit line increase or cross-selling a related service and by providing new services, including voice portals, voice activated dialing and e-mail reading.

For both enterprises and their customers, the financial benefits of a properly implemented speech recognition system are huge whether a speech system is new or added to an existing Interactive Voice Recognition (IVR) application.

The benefits of speech self-service extend far beyond the measurable financials that are required to obtain investment approval. Enterprises that implement speech systems should expect to see improvements in employee satisfaction, as agent burnout and turnover decrease and repetitive inquiries are offloaded to automated systems. Companies will also see an improvement in customer satisfaction, particularly if the company has speech-enabled a touch-tone based IVR application and to reduce the cost of transaction in call center than a legacy system.

The analysis will highlight the naturalness of Natural Language (NL) speech applications move into the mainstream, chances are the enterprise will install the technology during the next 12-18 months. Kai-Fu (2004), so that the enterprise will enjoy the 85% of waste control and 50% of cost reduction in call centers.

REFERENCES

Berger, A.L., S.A.D. Pietra and V.J.D. Pietra, 1996. A maximum entropy approach to natural language processing. Computat. Linguistics, 22(1).

Bill, G., 2004. Chairman and Chief Architect, Microsoft Corporation. Microsoft Speech Server VISION-Introduction CD-ROM-Part No. 098-97420-2004.

Chu-Carroll, J. and B. Carpenter, 1999. Vector-based natural language call routing. Comput. Linguistics, 25(3): 361-388.

Fugate, L.S., 2013. Contact Center Business Applications. Publisher, Speech Technology Magazine, May-June 2013.

Kai-Fu, L., 2004. Vice president, speech technologies, Microsoft Corporation. Microsoft Speech Server VISION-Introduction CD-ROM Part No098-97420-2004.

Manoharan, R., K. Vivekanandan and V. Sundaram, 2009. A software agent for speech abiding systems. J. Comput. Sci., 5(2): 90-96.

Povey, D., B. Kingsbury, L. Mangu, G. Saon, H. Soltau and G. Zweig, 2005. fMPE: Discriminatively trained features for speech recognition. Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05), pp: 961-964.

Soltau, H., B. Kingsbury, L. Mangu, D. Povey, G. Saon and G. Zweig, 2005. The IBM 2004 conversational telephony system for rich transcription. Proc. IEEE ICASSP, 1: 205-208.

Zach Epstein, 2011. Retrieved form: http://www. bgr. com/2011/05/05/major mobile-app-store-revenue-will-grow-77-7-in-2011/.