IJETAE
EXPLORING RESEARCH AND INNOVATIONS

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 1, Issue 2, December 2011)

Development of *iSpeak*: A Voice-activated Relationship Management System

Aderemi A. Atayero, Adeyemi S. Alatishe, Juliet O. Iruemi

Department of Electrical and Information Engineering, Covenant University, Nigeria

atayero@ieee.org

***Abstract*—A constant source of frustration for subscribers of mobile telephony in Nigeria is the quality of customer care service. The ubiquitous IVR systems deployed by service providers often ends in long and winding texting of digits that terminate in calls to agents with poor CRM attitudes. Automation of most of the functions of the human agent goes a long way in mitigating this problem. This paper describes *iSpeak* – a system designed to reduce the human–to–human (H2H) interaction in the complaint-lodging and solution provision process to a minimal level where it is not possible to eradicate it totally by a replacement with human–to–system (H2S) interactivity. *iSpeak* has an inherent capacity for improving the efficiency and drastically cutting CRM cost of corporate organizations. This comes with the attendant advantage of improved business-customer relationship.**

Keywords* – *Automatic Speech Recognition, Customer Care Service, Speech-control, Customer Voice Model, Voice Print, Voice Recognition

I. INTRODUCTION

The need for an automated system that minimizes the human presence and consequently the *human factor*, in customer relations cannot be over-emphasized. This is especially so in the telecommunications sector of developing economies such as currently exists in most sub-Saharan African nations with the advent of the GSM technology. *ISpeak* is a system developed for the specific purpose of automating the complaint lodging and solution provision through troubleshooting with the aid of an automated speech-activated relationship management system. It is imbued with inherent capacity for improving the efficiency and drastically cutting cost of Customer Relationship Management (CRM) departments of corporate organizations.

The rest of the paper is briefly summarized as follows. Section II gives a brief description of the system, its component parts, and the methodology of operation together with explanation of the system peculiarities. The modeling of the system is discussed in section III. This is done with the aid of a diagram explaining the abstract description of the system.

The platform for system development and a flow chart giving detailed sequence of system operation is presented. Section IV explains the deployment scenario with the aid of a diagram and section V gives a general conclusion.

II. SYSTEM DESCRIPTION

Customer calls in either to lodge a complaint or access the Frequently Asked Question Dataset (FAQ DS). The system automatically generates a Voice Print (also called voice template or voice model) using the statistical characteristics of the speaker's voice [1] and compares it with the Customer Voice Model (CVM) in the CVM dataset. If this particular customer's CVM is in the dataset, the system personalizes the session by retrieving relevant information about the customer. A personalized greeting is used to verify the correctness of customer recognition. On the other hand, if the CVM was not recognized in the CVM dataset, it is automatically added and the dataset is updated. A message is promptly generated to notify the customer that he is about to be added to the company's database to make subsequent transactions faster. All necessary information about the new customer is obtained and used to update the CVM dataset. The more information the system gets from the customer, the better for the Automatic Speech Recognition (ASR) module. This information is used to generate a Statistical Language Model (SLM) (a statistical grammar generating method used for speech recognition) for the particular customer. The FAQ DS is simultaneously updated with the registered complaint or comment and a possible solution is proffered.

If no such solution exists (this being the first time such a complaint is filed), the system tries to generate a new solution by inference from existing solutions using a knowledge-based technique. This newly inferred solution is then used to update the Solution Dataset (SDS). If it fails to generate a satisfactory solution, the call is routed to the Complaint Dataset (CDS) where a solution is later proffered by the Service Provider Diagnostic Expert (SPDE). The new solution proffered is used to update the SDS for future use.

This is the only instance when a human operator is physically engaged in the work of the system. The involvement of the SPDE is transparent to the customer. To achieve the aim of eliminating actual human contact, the FAQ DS is carefully, extensively and exhaustively built with a corresponding SDS that takes the customer through a *stepwise* troubleshooting and resolution process.

The Voice User Interface (VUI), which is the voice analogue of the Graphic User Interface (GUI) is designed to be maximally user friendly, without ambiguity and just enough redundancy to collect enough voice data for generating the customer's SLM without causing irritation [2].

It is a recognized fact in the ASR industry that most people prefer to "*Recognize*" rather than "*Generate*" i.e. think-up answers. For this singular reason, a verbal menu that requires the customer to punch a key, give a voice command or merely repeat the word when the desired option is mentioned is adopted. The last option known as "*barge-in*" can increase the feeling of interactivity on the part of the customer, while simultaneously reducing the time spent in the overall process. The *iSpeak* system reduces the human presence in the call center to a minimum. This is desired to cut cost, reduce response time and most importantly, eliminate other "negative" human related aspects of customer relations. The Dialogue Logic (a.k.a. call flow) is designed to determine how the system responds to what the caller/customer has just said and asking the appropriate follow-up questions or reading out information from the *iSpeak* Database (*iSpeak* DB), which is made up of several Datasets. For the smooth implementation of this process, a Call Flow Dataset (CFDS) is generated from the onset and subsequently updated as the system grows. The *iSpeak* system is a dynamic system that constantly grows and perfects itself with use.

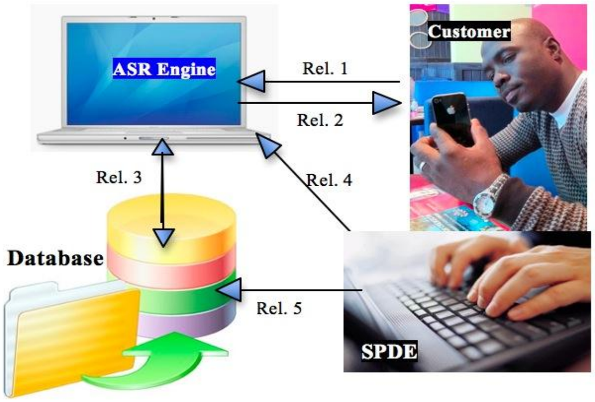
The peculiar properties of the system are described below:

- Every incoming call is logged into a separate dataset from which the system can retrieve additional information as the need may later arise using data mining techniques.
- The ASR module automatically updates its dictionary (a list of words and their pronunciations) after each session to accommodate new prosody (intonation, stress pattern, and rhythm of speech), which helps in creating more than one possible pronunciation for each word. This automatically improves the ASR module of the system with every new call.
- The endpoints are so selected as to accommodate customer's reflection time. This takes into consideration the fact that most customers are not used to *iSpeak* and may need some time to respond accordingly. This process is however eased by using the *recognize* mode for dialogue prompts.

The system gets better with each new call for each particular caller. A new call will increase the customer FAQ DS and SDS as well as improve the ASR module by the addition of new prosody. Each time an old customer (i.e. one with his information in the database) calls in, his/her records are updated and the voice model is improved upon for better recognition on the next call. This subsequently leads to the generation of a better statistical language model for a particular customer, which in turn increases the efficiency of the ASR module. The system also captures Non-Verbal Audio (NVA), which includes all audio outputs from a spoken language system that is not speech but mannerism (e.g. umm, ah, oh) that convey a meaning. These audio outputs are used to improve the voice model of the particular customer [3].

III. *iSPEAK* SYSTEM MODEL

The four basic components (i.e. entities or objects) used to model the system are: *iSpeak*, Customer, *iSpeak* DB and DA. The provider's computer hosts the *iSpeak* DB. The SPDУ is a domain expert in charge of complaint resolution. The customer supplies the necessary information that is fed into the database at the enrolment stage. Finally, the system generates new solutions from the existing, using a knowledge-based technique such as fuzzy logic. The system also collates solutions and mails them to customers.



Rel. 1: lodgeCompliant, makeEnquiry;
Rel. 2: generateVoicePrint, personalizeCVM, mailSolution;
Rel. 3: addCVM, mineData, generateNewSolution;
Rel. 4: checkUnresolvedCompliant, profferSolution;
Rel. 5: useDatabase

Figure 1: Abstract Description of *iSpeak*

Figure 1 is a schematic diagram showing the replacement of a H2H interaction model by a more resilient H2S interaction model. This is another potential benefit of our system. It serves to cut the front desk/call office administration costs, which as studies show can be quite high [4].

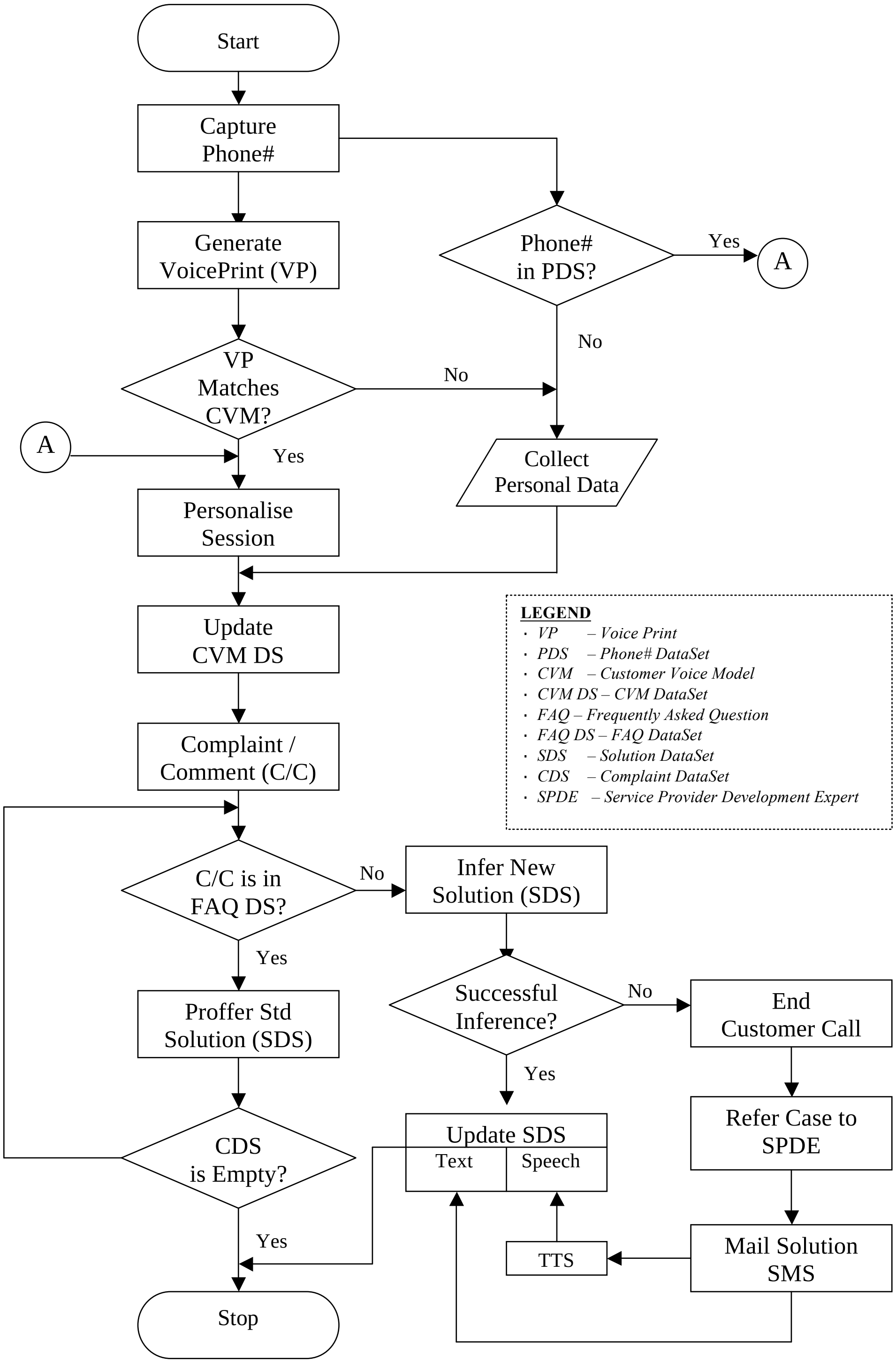


Figure 2. Operational Flow Chart for *iSpeak*

Furthermore, the feature reduces response time and eliminates the possibility of additional conflicts that might have emanated in the course of complaint-resolution in the H2H model. It is important to note that the *iSpeak* is a "Self-starter" because it can initiate dialogue with a customer by mailing solution through a text message (or via e-mail). This takes care of the earlier lodged complaints without a ready solution in the SDS.

A. System Development

Figure 2 gives a detailed description of the system in the form of a flowchart. The system is implemented using the VoiceXML capabilities of Object-Oriented programming packages such as Borland C++ Builder, Microsoft Visual Studio.Net and SALT based Microsoft Speech Application SDK (SASDK) [4].

IV. DEPLOYMENT ARCHITECTURE

The deployment architecture is shown in Figure 3 [5]. The recommended platform for deployment is a telephony layer implementing Cisco Voice over IP (VoIP) for providing interconnection between the Public Switched Telephone Network (PSTN) and Voice Gateway servers providing call processing, control and call flow execution. This platform has been tested and proven [4]. The functions of the elements of the deployment scenario are described below:

a) Media Gateway – Provides the signaling and media conversion between the PSTN and the IP-based network converts PSTN signaling protocols to IP-based Session Initiation Protocol (SIP) and its media formats to IP-based Real-time Transport Protocol (RTP) [6].

b) Service Controller – Provides SIP-based routing and load balancing between SIP-enabled devices.

c) VoiceXML Browser – This is the voice-enabled service creation environment. It executes applications developed according to the VoiceXML specification.

d) ASR Server – Provides automatic speech recognition service.

e) TTS Server – Provides Text-to-Speech (TTS) service. It converts a text string to audio signal, which is streamed through RTP to Media Gateway. It uses the SIP registration method to notify the service controller of its availability. The VoiceXML Browser communicates with the TTS Server when it needs to convert text to audio for onward delivery to the user [7].

f) Web Server – Standard HTTP server that hosts the voice site content such as VoiceXML scripts, audio prompts and grammars.

g) Manage/Monitor Voice Site Content – Provides a component server and content management view. This affords the system administrator the opportunity to update content, start, stop or/and check component server status.

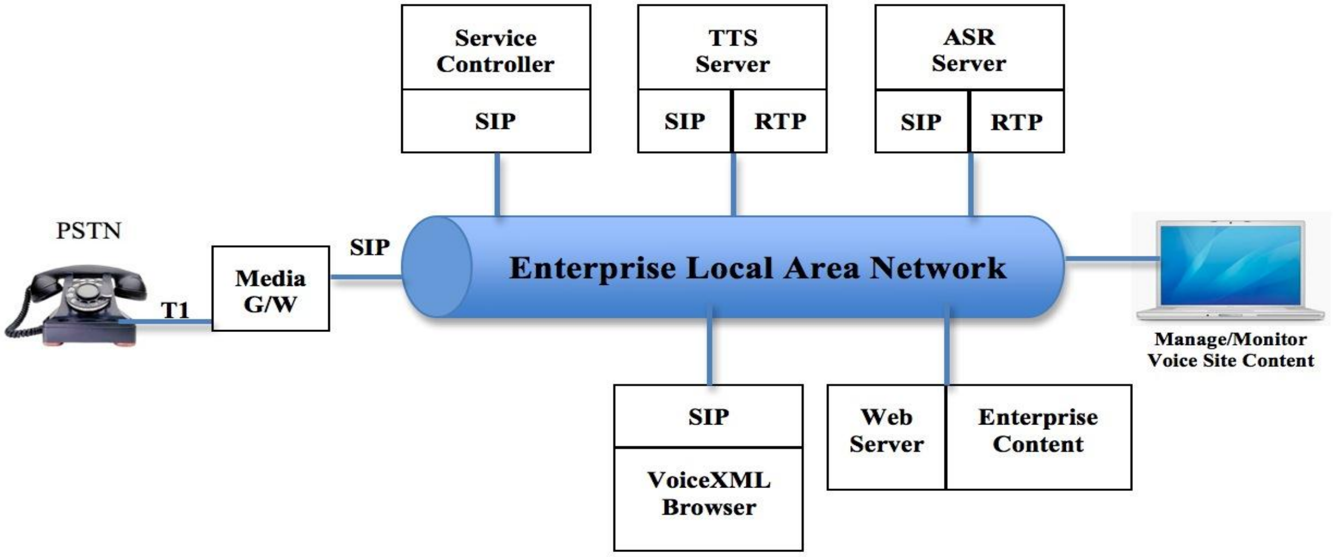


Figure 3. *iSpeak* Deployment Architecture

V. Conclusion

This paper has described the Design, Development and Deployment of a voice-activated Relationship Management System, which eradicates the need for a human operator, reduces the budget allocation of corporate bodies and most importantly, improves the B2C relationship, which is often damaged by inevitable flaws of the human character. We believe that the system will go a long way in improving the iSpeak quality of the Nigerian corporate environment in particular (and any corporate environment in general). The system we propose is dual deployable (both on the standard PSTN and on the VoIP network) and as such promises to be quite versatile. The system is profitable in any organization seeking to cut cost in its CRM budget, while simultaneously improving customer relationship.

References

[1] A.A. Atayero, “Comparative Analysis of Statistical Characteristics of Speech Signal” ISBN 5-85124-133-0, pages 69-73, Moscow, 1999.

[2] P. Giangola and Jennifer Balogh, “Voice User Interface Design”, Addison-Wesley.

[3] C.H. Lee, et al., “A study on natural language call routing”, Proceedings of IEEE 4th workshop on Interactive Voice Technology for Telecommunications Applications, IVTTA '98, pp.37-42, Sep. 1998, Torino, Italy.

[4] Speech Technology magazine, Jan/Feb. 2004.

[5] Brian Marquette, “Voice-Enabled Applications Deployed Using the Component Server Architecture”, 2001 SandCherry, Inc.

[6] H. Schulzrine, J. Rosenberg, “The Session Initiation Protocol: Internet-centric signaling”, IEEE Communications Magazine, Vol. 38, Issue 10, pp.134-141, Aug.2002.

[7] M. Bacchiani, et al, “Deploying GOOG-411: Early lessons in data, measurement, and testing”, Proceedings ICASSP, pp.5260-5263, Mar.31-Apr.4 2008, Las Vegas, NV, USA.