http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 13 (12): 1953-1960, 2014
ISSN 1812-5638 / DOI: 10.3923/itj.2014.1953.1960

Speech Enhancement Using an Artificial Bandwidth Extension Algorithm in Multicast Conferencing Through Cloud Services

[1]G. Gandhimathi and [2]S. Jayakumar
[1]Department of Electronics and Communication Engineering, India
[2]School of Computing Science and Engg, Periyar Maniammai University,
Vallam-613403, TamiNadu, India

Abstract: In the emerging scenario, Multicast conferencing plays an important role in e-education and e-business The multi conferencing systems have several shortcomings such as the conventional select-and-forward conference bridges gives inconsistent speech quality, due to full mesh and multicast architectures needs a large bandwidth at the endpoints. For a large scale voice conferencing through IP, now-a-days a tandem free architecture is coming up to reduce the bandwidth since it uses the compressed Low bit rate (Narrow Band) speech. The speech quality is to be increased by employing Artificial Bandwidth Extension (BWE) algorithm provided by Application Software Provider (ASP) at the endpoints. In this study, a new approach, the SaaS (Software as a Service) of cloud computing service is introduced to import Artificial BWE algorithm in to the cloud and the performance of SaaS over ASP is analyzed.

Key words: Multicast conferencing, wide band speech, speech codec, SaaS, artificial band width extension

INTRODUCTION

Wide band speech: Today most of the speech coding systems are based upon narrowband speech, nominally limited to 200-3400 Hz with 8 kHz of sample rate. The intrinsic bandwidth limitations in the Public Switched Telephone Network (PSTN) enforce a limit on communication quality. The increasing access of the end-to-end digital networks, such as the second and third generation (2G and 3G) wireless systems, ISDN and Voice over Packet networks (VoIP). P.J. Smith compared the quality of the speech over conventional VoIP conference with tandem connections to that of a Tandem free Operation (TFO) conference and proved TFO-MS/I system is better than the conventional VoIP bridges (Smith, 2002).

Multicast Voice Conferencing can use wider speech bandwidth that will be offered communication quality and creates the feeling of face-to-face communication. Normally the energy level is more below 7 kHz of the speech signal, only less energy may extend to higher frequencies and mainly on unvoiced sounds. For wideband speech coding, the signal has to be sampled at the rate of 16 kHz. The study is formulated as, the cloud computing is to be discussed in cloud, the voice conferencing through PSDN and ISDN are to be discussed in related work and the artificial BWE algorithm for multicast conferencing in cloud services is described in methodology. The deployment of ASP is discussed in conclusion.

Cloud

Cloud computing: The term Cloud computing is the rescue of computing as a service not a product, there by shared information, software and resources are provided to computers and other devices as a metered service through an Internet. The entire IT industry can be moved towards to the cloud computing. Developers with modern, innovative ideas for new Internet services need not require the large initial capital in hardware to deploy their internet service or the human expense to operate it. Figure 1 shows the various cloud models existing in the world, the cloud models is to be adopted by the organization depend upon their nature and each model can able to provide its software, platform or its infrastructure as the service (SaaS, PaaS, IaaS) by Pay as you use method as illustrated in Fig. 2.

SaaS: Software as a service can be characterized as "deployment of Software in the hosted service and it can be accessed over the Internet instead of installation and maintance". The only requirement for SaaS is a computer along with a browser as shown in Fig. 3. SaaS is a frequent subscription based model delivered to the end

Corresponding Author: G. Gandhimathi, Department of Electronics and Communication Engineering, India

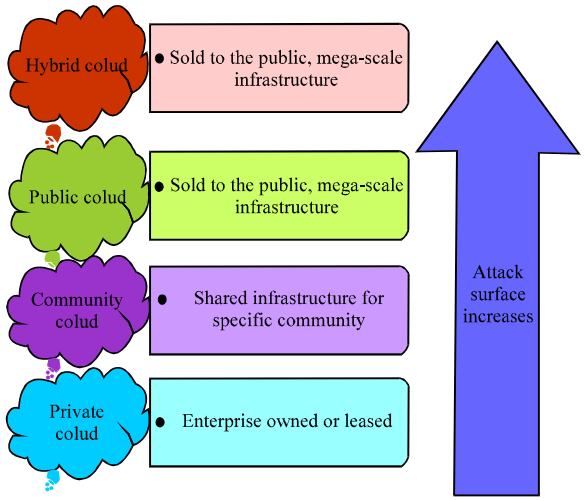

Fig. 1: Cloud models

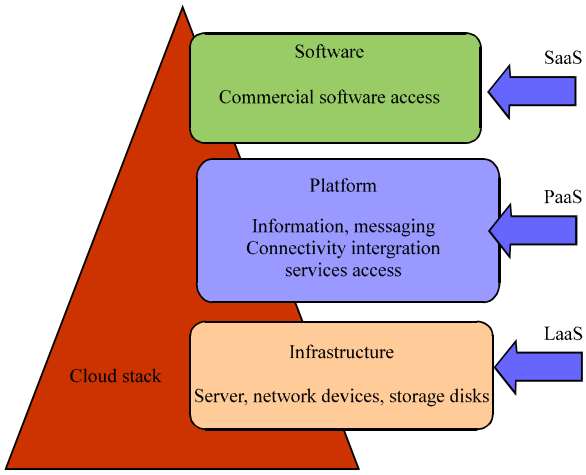

Fig. 2: Cloud services

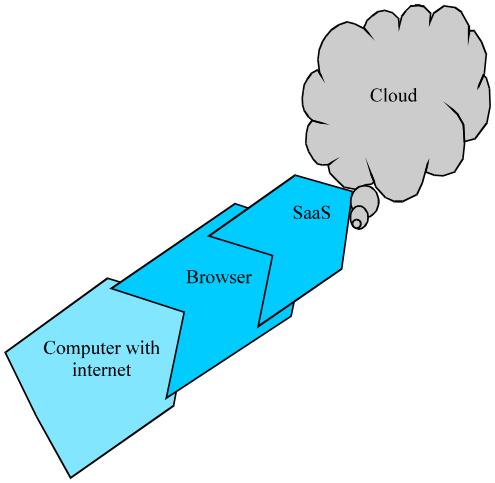

Fig. 3: Requirement of SaaS

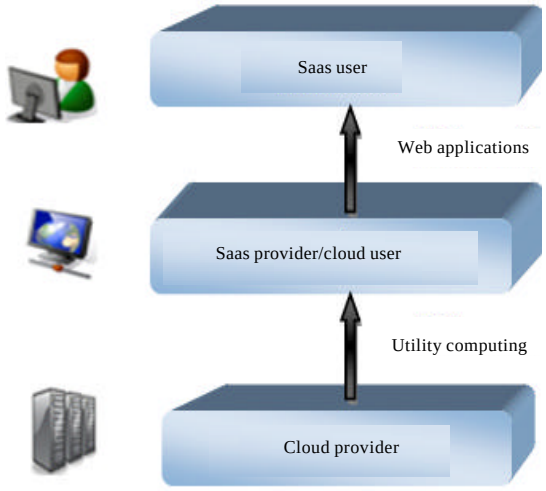

Fig. 4: SaaS Stack

user on demand. The Customer Relationship Management (CRM) solutions provides the SaaS. The applications are run on the SaaS provider's servers. The provider manages access to the application, including availability, security and performance. SaaS users have no software or hardware to buy, install, maintain, or update and can access the applications easily as shown in Fig. 4.

Categories of SaaS: There are two major categories of SaaS:

- Line-of-business services, accessible to the organizations of all sizes and these services are sold typically to customers on a subscription-basis
- Consumer-oriented services, accessible to the general public that are often provided to consumers at no cost and are managed by advertising

Lot of benefits are (Fig. 5) there for both the SaaS providers and the end user. The End user can utilize these software from a third party SaaS vendor (Jurcicek *et al*., 2011; David, 2011). They need not spent their budget on software services, moreover applications can be delivered over the Web which enables the end user to extend the life-cycle significantly by updated desktop technology. On the other hand, from a SaaS provider perspective, if the SaaS applications is scalable, more end users are added, the provider will able to develop multi-tenancy as a core competency, that may lead to provide high-quality offer to the end user at low cost.

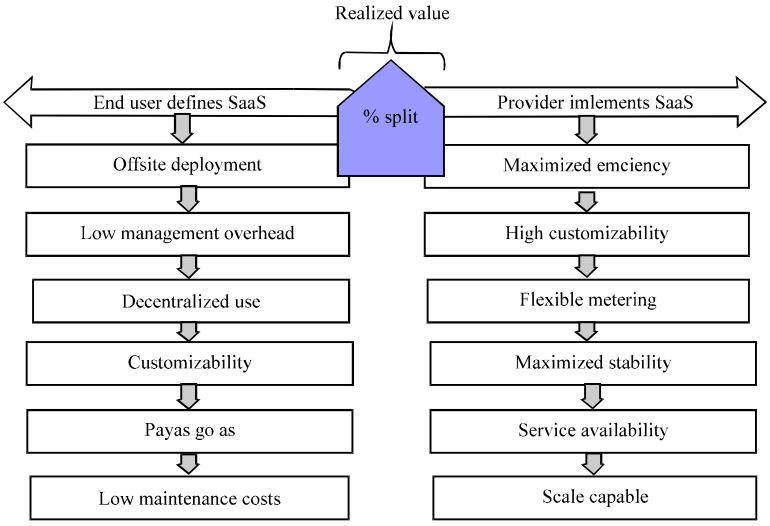

Fig. 5: Tradeoff between SaaS provider and end user

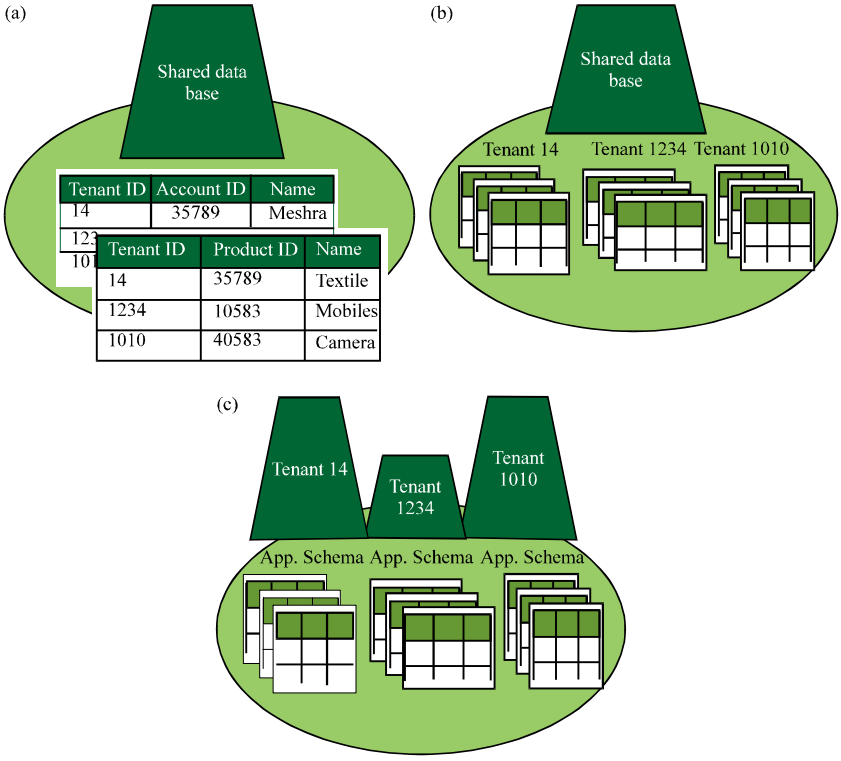

Fig. 6(a-c): (a) Shared Database, Shared Schema, (b) Shared Database, Separate Schema (c) Separate Database

Database schema: Ad hoc/Custom (Pseudo SaaS), Configurable (Quasi SaaS), Configurable, multi-tenant-efficient (Semi SaaS), Scalable, configurable, multi-tenant-efficient (True SaaS) are the four ways (maturity models) to host the applications on the SaaS architecture. For implementation of SaaS, the data schema can be varied, as shown in Fig. 6-c based on the maturity models selected.

LITERATURE REVIEW

Speech codecs: In standard narrowband VoIP calls, the narrowband codec like G.711 offers the voice signal is sampled at 8 Khz, that provides an effective voice pass-band of about 200 to 3,400 Hz and has the synthesized speech. Recent VoIP applications, the wideband voice codec like G.722.2 (Adaptive Multi Rate-Wide Band) offer 16 kHz sample rate, that provides an effective pass-band of 50 to 7,000 Hz wideband voice through IP phone (referred from mobistar) has much higher fidelity voice call, more or like to talk someone directly in the same room rather than over a phone and more DSP horsepower is needed (4 to 5 DSP cycles than G.711). In summary the use of G711 vs AMR-WB can seem right if battery usage against bandwidth in the air interface trade-off is constructive (Brent, 2007).

During transmission, in addition to the actual speech coding of the signal it is necessary to use channel coding, to avoid losses due to transmission errors. Normally, speech coding and channel coding methods have to be preferred in pairs, since the more important bits in the speech data stream confined by channel coding, bit rate scalable speech codec is used to get the excellent overall coding results for IP telephony and IP broadcast applications (referred from service.research.com)

Artificial Bandwidth Extension algorithm: High quality speech can be achieved is by applying a Wideband (WB) coding scheme. But this solution requires an expensive network upgrade. A possible solution is to artificially extend the NB speech signal to High-band (HB) frequencies from 3.4 kHz to 7 kHz (Aarts *et al*., 2003). This technique is apparent to the transmitting network, as it will be implemented only at the receiving end. Artificial BWE algorithm (ABWE) has to be implemented in speech and audio compression applications (Jax and Vary, 2003). They are based on code book methods or linear mapping or statistical mapping (Katsir *et al*., 2011). In these methods, the low frequency band of the spectrum is encoded with an existing codec and the frequency band is roughly parameterized using fewer parameters on ideal condition. The WB signal is predicted from the correlation between NB and WB signal. The block diagram is shown in Fig. 7. Yet, in multicast conferencing, the speech may come a variety of environments, the Mel Frequency Cepstral Coefficients (MFCC) based BWE algorithm can able to perform robustly even in noisy environments (Seltzer *et al*., 2005; Cong and Hu, 2010; Katsir *et al*., 2011).

Mathematical model for extracting MFCCs from WB and NB speech: By using Short time Fourier Transform, the speech signal is segmented in to number of frames. Define $|Z|^2$ is the power spectrum of a frame, then the generated MFCCs for the future extraction process is modeled as:

$$Z = C\log(W|Z|^2) \quad (1)$$

where, W is the weighting coefficients matrix of mel filter Bank and, C is the Discrete Cosine Transform Matrix (DCT). For simulation by MFCC method the MATLAB software (ASP) is used and the wave file of sentence 'seed feed seed feed seed ' spoken by the male speaker over 10 sec is taken. The results in time domain and its spectrums are shown in Fig. 8. For calculating the Power spectral density under noisy environment, the Multiple Signal Classification algorithm (MUSIC) based on subspace method (Gandhimathi and Jayakumar, 2012) with N = 1024 FFT points are taken and it is shown in Fig. 9.

Mathematical model for MUSIC method:

$$\hat{P}_{MUSIC}(f) = \frac{1}{e(f)H\left(\sum_{K=p+1}^{N} Vk\ VkH\right)e(f)} \quad (2)$$

$$= \frac{1}{\left(\sum_{K=p+1}^{N} |VkH)e(f)|\right.} \quad (3)$$

$$\text{The Eigen vector } e(f) = [\,1\ e\ (j2\pi f)\ e\ (j2\pi f.2)\ e\ (j2\pi f.4)\ldots\ (j2\pi f(N-1))\,] \quad (4)$$

VoIP: Bandwidth optimization and codec latency: From the above over view the low-bandwidth codecs are quite efficient. For example, G.729 will compress 10 milliseconds

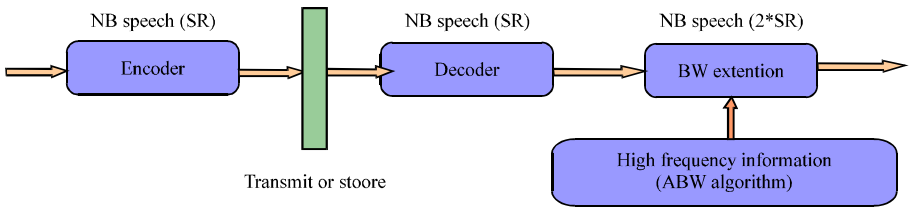

Fig. 7: Block diagram of artificial bandwidth extension of speech signal

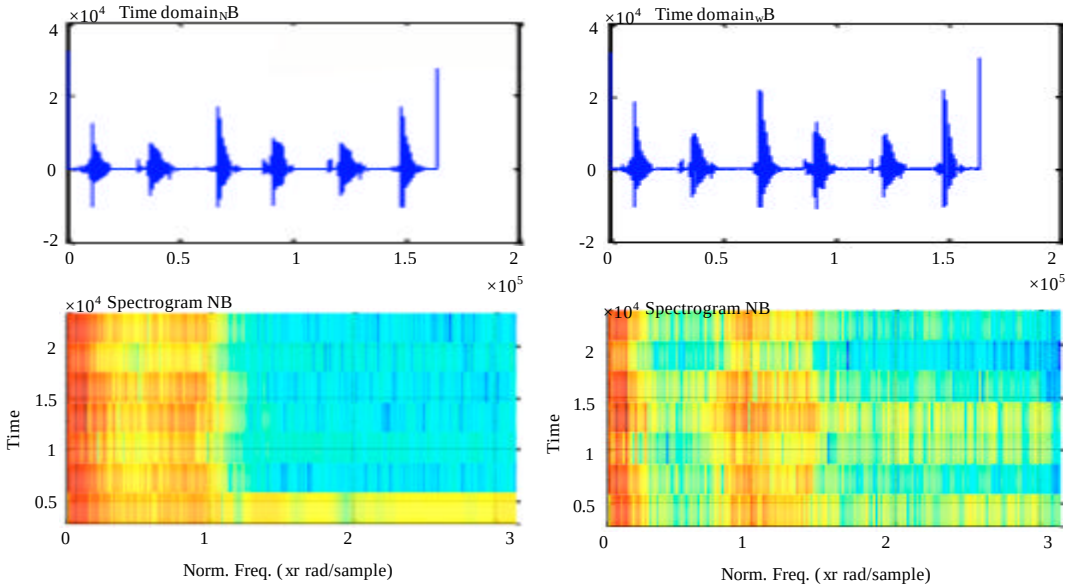

Fig. 8: Time domain and frequency domain of NB and WB Speech signal

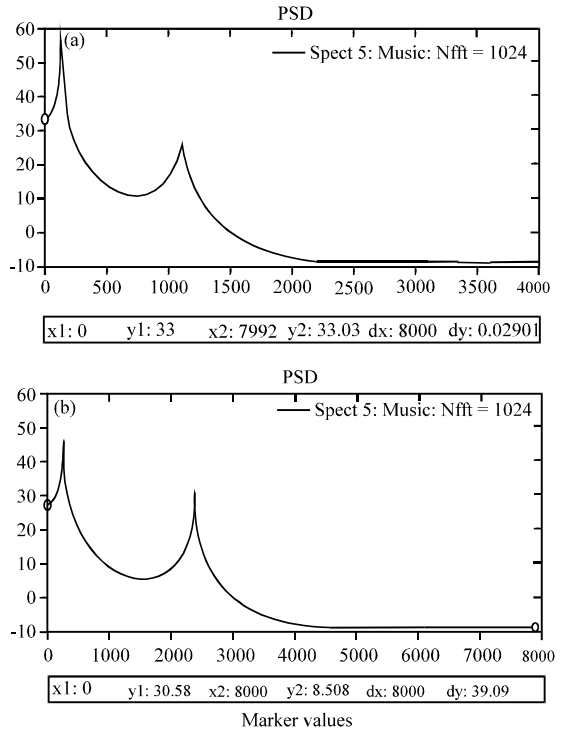

Fig. 9: PSD of NB and WB Speech signal

of audio to 10 bytes and G.723.1 encodes 30 msec frames to 24 or 20 bytes. The IP, UDP and RTP header overhead

Table 1: Over head of speech codecs over VoIP

Codec	Nominal bit rate(kbit/s)	Frame length(ms)	Frame size (bytes)	Packet over head	Actual bit rate (kbit/s)
G.723.1	6.4	30	24	167%	17.0
G.723.1	5.3	30	20	200%	16.0
G.729	8.0	10*3	10*3	133%	18.6
GSM 06.10	13.0	20	33	121%	29.2

have to be considered. Since, the compressed audio frames are sending over VoIP on Real Time. Overall 40 bytes overhead is needed per pocket. This is noteworthy when compared with the size of the compressed audio frame. The Table 1 illustrates the overhead for several low bit rate codec. For calculation, one frame / pocket is taken for G723.1 and GSM and three frames/pocket is taken for G.729, since it works with 10 ms frame size. The bit rate values are to be calculated during only one direction of the call.

Adoptive play out algorithm: Der, introduced an adoptive play out algorithm for Real Time VoIP based on normalized least mean square algorithm. Later on he developed the enhanced bi-nominal play out algorithm, shown in Fig. 10 which in turn reduces both average delay and pocket loss rate (Der *et al*., 2003).

Codec enhancements: One of the Play out Controller is to be implemented at the receiver, that buffers the variable delay or jitter in the network and to give a fixed stream of

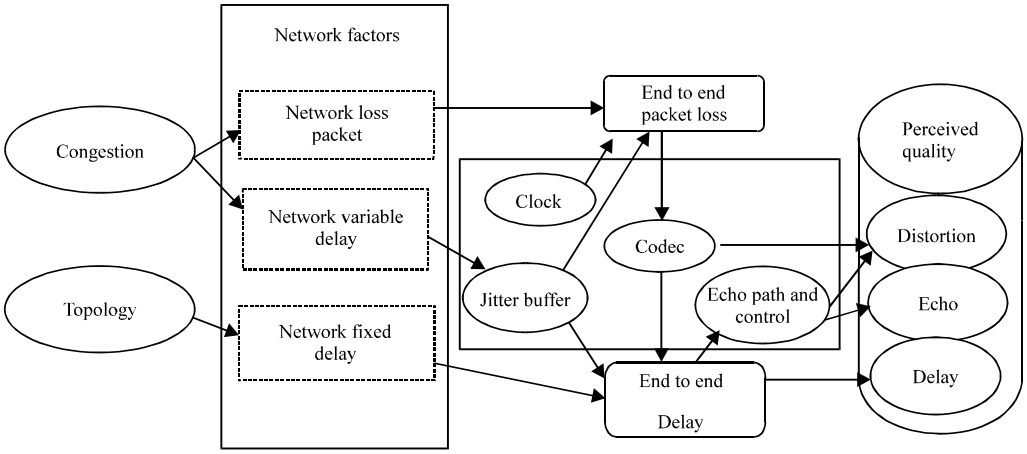

Fig. 10: Block diagram of playout algorithm

packets for the codec. At the receiving end the play out controller controls the size of the buffer, therefore, trades off packet loss and overall delay or latency. The design and algorithms for play out controllers have become much more refined in their ability to reduce additional delay and also screen the packet loss. Furthermore the play out controller can decrease the upshot of clock skew which can take place when the sender and receiver not being synchronized.

Codec independent packet loss (PL) concealment: There are two simple approaches for packet loss concealment. The first method, Zero Stuffing (ZS) is obtained by simply a lost packet is replaced with a period of silence of the same duration. The second method, is Packet Repetition (PR), here, the difference between two consecutive speech frames is assumed to be quite small. Hence, the lost packet is replaced by the previous packet. For example in practice, though, even a small change in, the pitch frequency is detected easily by the human ear. In addition, with this approach. It is virtually impossible to get smooth transitions between the packets. However, this approach performs quite well for very small probabilities (less than 3%) of packet loss. Recently, the ITU standardized a method for packet loss concealment in G.711, the codecs based on CELP have own their built-in packet loss concealment algorithm. This gives a reasonable suppression during the loss, but they suffer from their packet inter-dependencies.

MATERIALS AND METHODS

Today Cloud-based VoIP services may not even enter into the conversation were it not for the any-to-any

Fig. 11: AT and T and cloud

network connectivity of Multi Protocol Label Switching (MPLS) technology. Entering the scene a decade ago as a way to link diverse networks. MPLS laid the groundwork for the virtually anywhere, anytime communication of today and the advanced capabilities of the future. By carrying voice and data on the same network, the global AT and T MPLS backbone makes it easier to extend converged voice and data solutions to a range of fixed and mobile devices from desktop PCS to smart phones. As a result, you can have the reach, agility and enhanced network security you need for voice communications and a speedy highway to the AT and T cloud (referred from voice transformation) as shown in Fig. 11.

AT and T voice builder: The AT and T voice builder provides the new tool, as the web service, implemented on the AT and T speech Mashup (Web app+speech) portal. The systems records speech and validates the user utterances, process them to built the Natural voice (WB) and provides a web service Application Provider Interface (API) to make the voice available in real time applications through a cloud based processing platform shown in

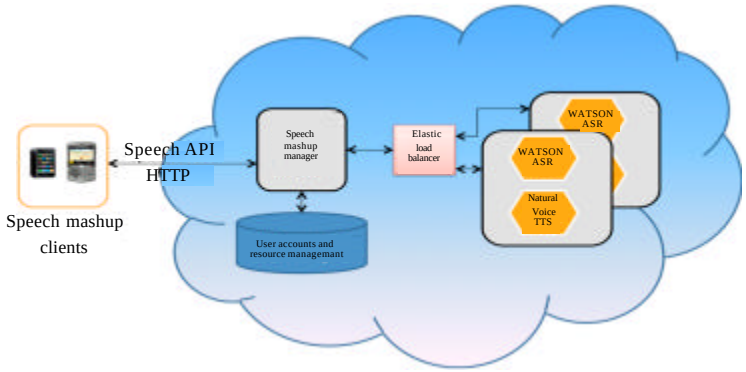

Fig. 12: speech mash up

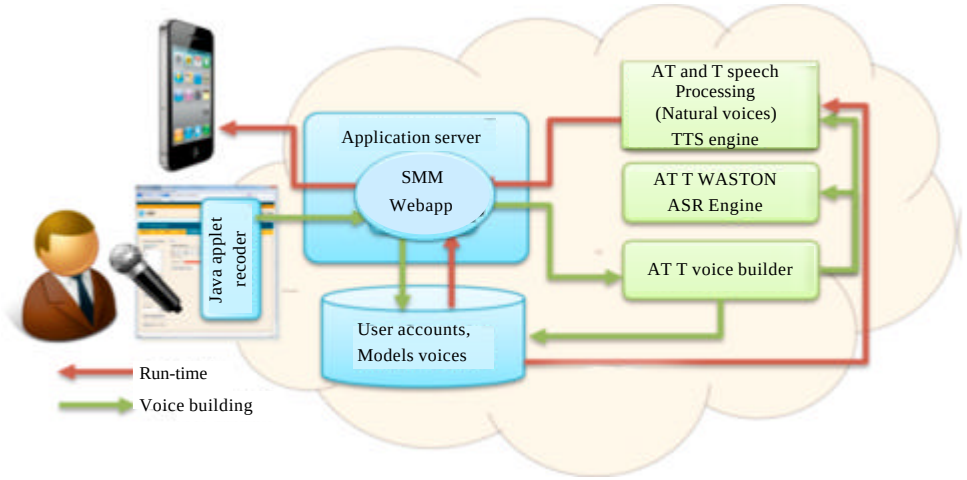

Fig. 13: Process diagram of conferencing in AT and T and cloud

Fig. 12. All the procedures are automated to avoid human intervention. It has commercial grade high quality speech enhancement system without the need to install speech processing software and equipment (Fig. 13).

Building a speech mashup for a mobile device:

- Register in the speech mashup portal for creating an account on AT and T servers
- Create a directory for the web application and related files (grammars, log files, etc.)
- Create and upload the grammars or using a built-in or shared grammar
- Build a speech mashup client using Java or JavaScript programming language
- Enable the client, enter in to the cloud

To minimize the number of round trips in the mobile network, the processing steps are tightly integrated and that reduce latency to get a better user experience.

End users access of SaaS applications: The steps followed by the end user can access the SaaS applications:

- Create end user account (for authentication) to access the account portal
- Numbers of subscribed applications are there in the Application page of the account portal. Select the desired application by clicking on the Lanch icon. Or
- The user can also access by means of Custom URL controlled by the organization that has been configured to access the application

CONCLUSION

Cloud computing and SaaS hands together would make a massive impact to the present way of working. Consuming and exposing the software services over internet would enable the organization and IT infrastructure to fully depend on it. Speech enhancement through cloud computing distribute any single application through the internet browser to more number of customers using a multi-tenant architecture. In addition from the customer perspective there is no upfront investment in servers or software licensing, on the other side the service provider, with just any one application to maintain, costs are low compared to conventional ASP.

REFERENCES

Aarts, R.M., E. Larsen and O. Ouweltjes, 2003. A unified approach to low- and high-frequency bandwidth extension. Proceedings of the 115th Audio Engineering Society, October 10-13, 2003, New York, USA.

Brent, L., 2007. Tip: Wideband vs. narrowband VoIP codecs. Texas Instruments. http://www.eetimes.com/design/signal-processing-dsp/4017506/Tip-Wideband-vs-narrowband-VoIP-codecs/

Cong, Z. and R.M. Hu, 2010. A novel mobile audio bandwidth extension algorithm oriented 3G communication. Proceedings of the International Conference on Multimedia Technology, October 29-31, 2010, Ningbo, pp: 1-4.

David, S., 2011. Cloud computing and crowd sourcing for speech processing: A perspective. SLTC Newsletter. Speech and Language Technical Committee, IEEE Signal Processing Society.

Der, R., P. Kabal and W.Y. Chan, 2003. An adaptive playout algorithm with delay spike detection for Real-time VoIP. Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering, Volume 2, May 4-7, 2003, Canadian, pp: 997-1000.

Gandhimathi, G. and S. Jayakumar, 2012. Efficient method of pitch estimation for speech signal using MATLAB. Special Issue IJCCT, 3: 107-111.

Jax, P. and P. Vary, 2003. On artificial bandwidth extension of telephone speech. Signal Process., 83: 1707-1719.

Jurcicek, F., S. Keizer, M. Gasic, F. Mairesse, B. Thomson, K. Yu and S. Young, 2011. Real user evaluation of spoken dialogue systems using amazon mechanical Turk. Proceedings of the12th Annual Conference of the International Speech Communication Association, August 27-31, 2011, Florence, Italy.

Katsir, I., I. Cohen and D. Malah, 2011. Speech bandwidth extension based on speech phonetic content and speaker vocal tract shape estimation. EUSIPCO, Barcelona, Spain. http://webee.technion.ac.il/Sites/People/IsraelCohen/Publications/EUSIPCO2011_Katsir.pdf

Seltzer, M.L., A. Acero and J. Droppo, 2005. Robust bandwidth extension of noise-corrupted narrowband speech. Proceeding of the INTERSPEECH 2005 Conference, September 4-8, 2005, Lisbon, Portugal, pp: 1509-1512.

Smith, P.J., 2002. Voice conferencing over IP networks. M.Sc. Thesis, McGill University, Montreal, Canada.