International Journal on Islamic Applications in Computer Science And Technology, Vol. 3, Issue 4, December 2015, 29-37

Speech-to-Speech Translation System for the two Holy cities

Syed Uzair Ahmed[1,a], Naveen N S Khan[2,b]

[1]Sir Syed University of Engineering and Technology, Karachi, Pakistan

[2]University of Hail, College of Computer Science and Engineering, Hail, K.S.A
[a]syeduzairahmed@yahoo.com, [b]naveen.skhan@gmail.com

ABSTRACT

A Speech-to-speech (S2S) translation system is proposed to translate Arabic speech to other languages and the main focus is on the holy cities of Makkah Al-Mukarramah and Al-Madīnah Al-Munawwarah (hurmain). They are amongst the world's most visited holy cities, specially every year only in Hajj (Pilgrimage) season more than 1.3 million international Mu'tamirs (pilgrims) visit the two holy cities. Moreover, throughout the year more than 5 million Mu'tamirs visit the holy cities. As the visitors are from all over the world speaking different languages but the Imams (Prayer leaders) in the Harmayn speak Arabic only during their Sermons. This paper is aimed at a Speech-to-Speech translator application mainly from Arabic to other languages in real time.
The idea proposed is for an application that could recognize the Voice of the speaker first, be able to recognize the language of the speaker and to ask for the language in which it needs to be translated. It recognizes spoken input, analyses and translates it, and finally utters the translation. It should pause and start depending on the voice of the person. Speech Translation is the process by which conversational spoken phrases are instantly translated and spoken aloud in a second language. Furthermore, this differs from phrase translation, which is where the system only translates a fixed and finite set of phrases that have been manually entered into the system. Any device could be used for translating it could be a mobile phone which could have an application, Laptop which could use software or any other Hand held PDA's.

***Keywords*: Speech-to-Speech, S2S, Sermon Translator, Translator.**

1. INTRODUCTION

The number of visitors to the two holy mosques (Masjid Al-Haram & Masjid Al-Nabwi) in Saudi Arabia is increasing every year, in 2012 the number of international visitors only for hajj reached up to 1.8 million and it is expected to increase immensely in the fore coming years after the expansion of the two holy mosques in Mecca and Medina (Inventory for Hajj, 1435). The visitors from all around the world speak and understand different languages. This paper proposes a S2S system from Arabic to different languages. Visitors come from all over the world speak and understand different languages other than Arabic. According to the statistics mentioned in (Inventory for Hajj, 1435) 15 to 20 percent of the visitors (both resident of KSA and visitors from offshore) are from Pakistan & India (Inventory for Hajj,

1435) who understand and speak Urdu, hence the paper describes real-time S2S translation from Arabic to Urdu. It can be later on implemented to any other language.

For real time and multilingual situation S2S translation is under the spot light and is one of the promising technologies as mentioned as one of the emerging technology in Gartner's hype cycle for emerging technology, 2014 (Hung, 2014). S2S comprises of natural language processing technologies namely Automatic Speech Recognition (ASR), Statistical Machine Translation (SMT), and text-to-speech (TTS) (Sangmi, 2015).

In recent years there have been evident improvements in S2S translation systems. Translation for Tactical Use (TRANSTAC), under the DARPA project; for instance IBM TRANSTAC S2S translation (Bowen, 2013), SRI IRAQCOMM™ and BBN Transtalk (Rohit, 2013). IBM, BBN Technologies and SRI have advanced in multilingual S2S translation and taken it to the extent of implementation on the mobile devices. These systems have been developed for Pashto, Dari, Farsi, Malay and Arabic (Rohit, 2013). The implementation of S2S on portable handheld devices have the challenges of high latency and low resources. TRANSTAC is aimed to address these issues for S2S translation in their latest developments by using the three basic systems: ASR, SMT and TTS (Bowen, 2013) (Rohit, 2013).

Even though, the recent studies stated above, address most of the issues for the S2S translation on mobile device but for Arabic to Urdu translation not much work has been done yet. This paper focuses on real time Arabic to Urdu S2S translation targeted at the Arabic Sermon in two Holy Mosques. This is a one way translation of the sermon (Arabic) to Urdu language that will be broadcasted to the users on their mobile devices through an application on their mobile devices. Figure 1 shows the proposed system block diagram.

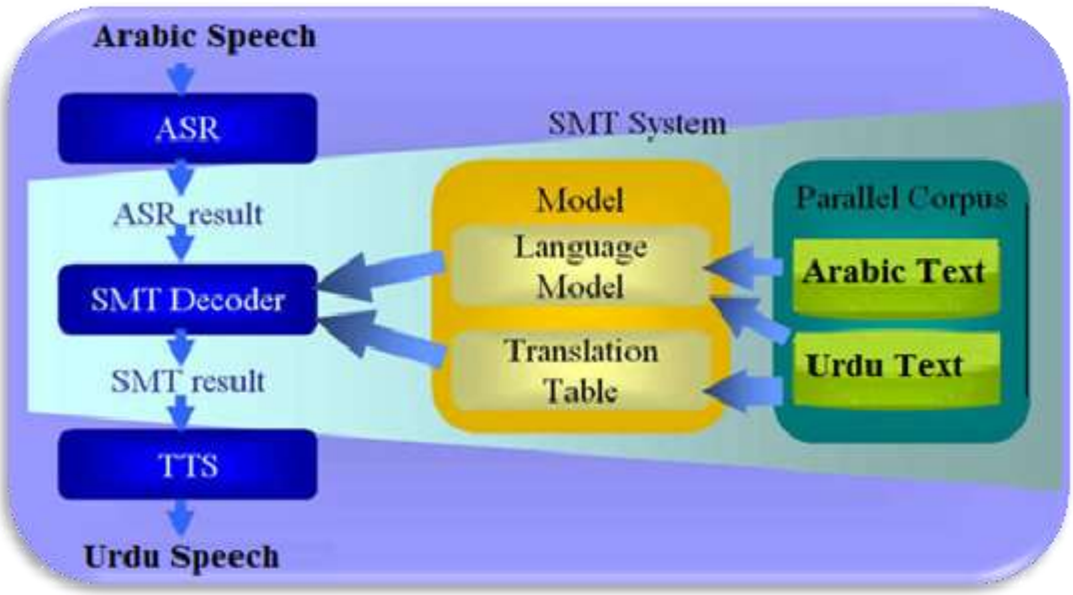

Figure 1. Proposed System Block Diagram

2. PROPOSED SYSTEM

2.1 Arabic Speech Corpus

Speech corpus is an important requirement for the development of S2S translation system. There is a vast Arabic corpus available (Mohammad, 2010), which had been developed by King Abdulaziz City of Science & technology (KACST). For the proposed solution the speech data from the Harmayn archived sermons and Quranic corpora can also be included. Through MATLAB programming the following tasks has been achieved (Mohammad, 2010).

- Segmentation of the Arabic Speech.
- Conversion of the speech data into its parameters
- Filenames convention
- Generation of transcription files for training and testing purpose

2.2 Automatic Speech Recognition (ASR)

ASR allows a computer to understand and identify whatever a speaker speaks in a microphone or a telephone. Arabic – Urdu S2S translation requires an ASR which can identify the speech of the *Imams* (Speakers) of the Holy Mosques, which are in Arabic to further process it before translating it to other languages; in this case Urdu. There are many open source ASR systems available such as HTK, ISIP, AVCSR and CMU Spihnx-4 (H.Satori, 2007). For the purpose of proposed S2S translation system CMU Spihnx-4 is considered to perform ASR duties, which is an open source system from Carnegie Mellon University (CMU) which works on Hidden Markov Model (HMM). In (Sangmi, 2015 ; H.Satori 2007 ; Mohammad, 2010) the CMU Spihnx-4 based on HMM was used and showed remarkable recognition rate on an average 95% correct recognition rate, irrespective of the speaker and sentences. Figure 2 shows the Architecture for CMU Spihnx-4, it shows the major components of CMU SPihnx-4 and the major Speech Recognition (SR) database components (Mohammad, 2010)

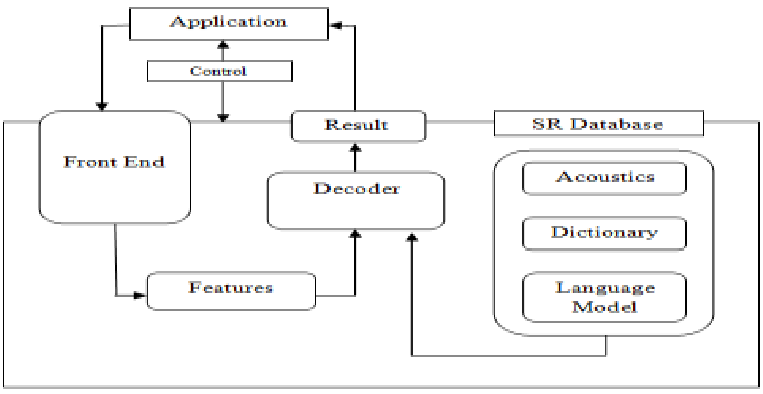

Figure 2. CMU Spihnx-4 Engine Architecture

2.2.1 Feature Extraction

To convert speech signal into feature vectors, vectors can be extracted by applying the well-known dominating feature extraction technique known as Mel-Frequency Ceptral Coefficients (MFCC) on the spoken utterances and then can be used for training and testing purposes. These feature vectors contains unique characteristics for each recorded utterance and then be used as an input SR database (Mohammad, 2010).

2.2.2 Phonetic Dictionary

In ASR systems the phonetic dictionary works as an intermediate link between the language model and acoustic model. For a given transcription, to automatically generate the phonetic dictionary, a rule-based approach is proposed. The details of these of the dictionary cab be found in (M.Ali, 2010). Figure 3 shows the sample of the generated phonetic dictionary (Mohammad, 2010).

امُ	E AE: L AE: M UH
آمِنٍ	E AE: M IH N IH N
آيَاتُ	E AE: Y AE: T UH
أَبَدَ	E AE B AE D AE
أَبِي	E AE B IY

Figure 3. Sample Arabic dictionary

2.2.3 Acoustic Model

In order to recognize speech, acoustic model provides the Arabic tri-phones to be used by HMMs. CMU Sphinx-4 supports acoustic modeling based on tri-phones, the function of perimetrization of the probability distributions of the state emission probabilities is achieved in CMU Sphinx-4 using Continuous Hidden Markov Models (CHMM). The key feature of tri-phone model is that it models individual phoneme while captures the neighboring right and left phonemes at the same time (Mohammad, 2010).

2.2.4 Language Model Training

The grammar used in the system is looked into by the language model. CMU-Cambridge Statistical Language Modeling toolkit is used for the development of the language model (Mohammad, 2010).

3. MACHINE TRANSLATION (MT)

As for the proposed S2S Translation solution, the MT would be only unidirectional; i.e. from Arabic to Urdu. There is no need for Urdu to Arabic translation. For the definition of sentence structure, it is considered that most of the sentences (speech) if from Quran and Hadith (Quran corpus; Hadith Corpus), huge corpora is readily available in both the languages. Similarly, Quranic and Hadith Urdu is easily available on many online sources with extensive corpora (Quranic Urdu). Furthermore, to make the corpora more affluent the archived sermons for the past several years can also be found on (Archived Sermons of Haramayn). Using the Statistical Modeling Translation (SMT) (Adam, 2008) and large parallel corpora in (Quran corpus; Hadith Corpus; Archived Sermons of Hharamayn), MT is proposed for S2S Translation for Arabic – Urdu. CMU SLM toolkit can be used for the purpose of MT. Figure 4. Illustrates the SMT model with an Arabic input and Urdu output.

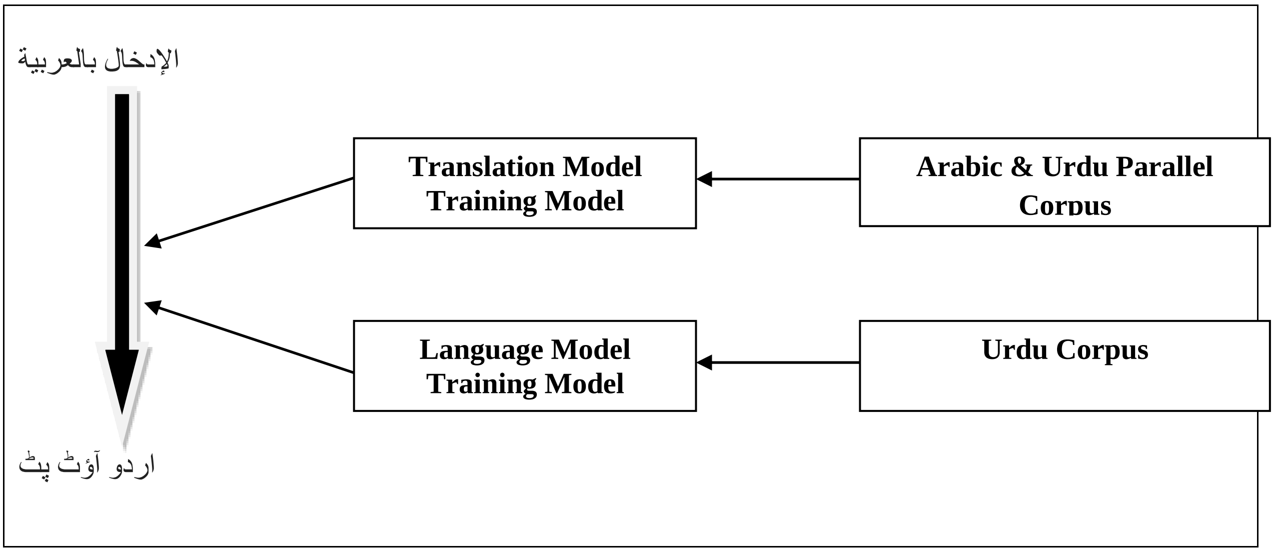

Figure 4. SMT Based Arabic to Urdu Translation

4. TEXT TO SPEECH SYNTHESIS (TTS)

The last part of the S2S translation for Arabic – Urdu comprises of converting the raw translated Urdu text to Urdu speech. TTS can be sub divided into two stages, first stage converts the raw text Input after certain algorithmic processing it converts the text into exact phonetic string that is to be spoken, annotated with prosodic markers. The phonetic representation is then processed to produce appropriate speech signal using a meticulous synthesis technique(Sarmad ,2004). Encompasses the Letter-to-Sound conversion for Urdu text-to-speech, details can be seen in (Sarmad ,2004). The TTS consist of a Natural Language Processor (NLP). The block diagram of the NLP for Urdu TTS is shown in the Figure 5.

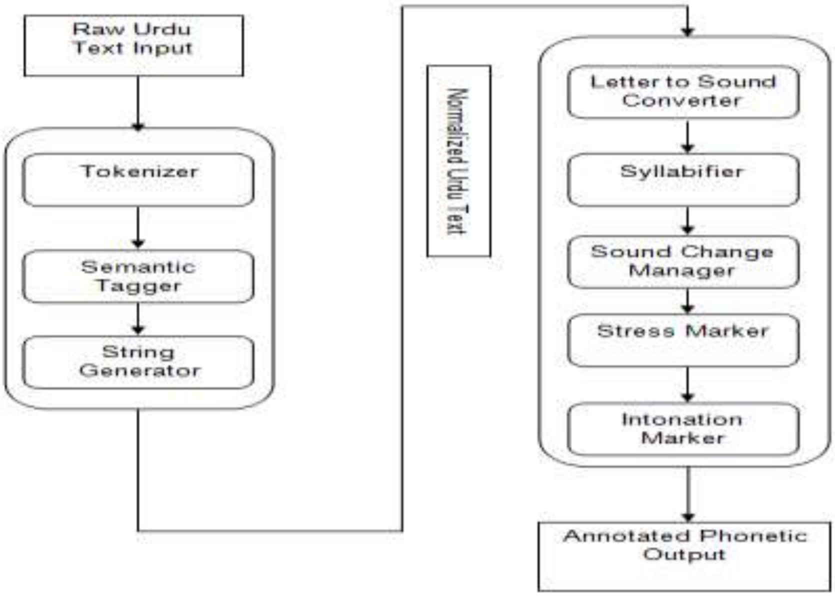

Figure 5. Architecture for Urdu TTS

5. PROPOSED ACCESS TECHNIQUE FOR USERS

Due to increased computational requirement, it is difficult to maintain low delay and latency for this real time S2S application on handheld devices like mobile phones and PDA's. Although there has been research going on for the enhanced, low latency and error free implementation of S2S techniques on mobile phones and handhelds. IBM's TRANSTAC and BBN's, Trans Talk have shown promising results (Bowen, 2013 ; Rohit, 2013); but still a lot more needs to be done. For the system proposed here, Figure 6 shows the architectural overview for the access technique for the users/listeners.

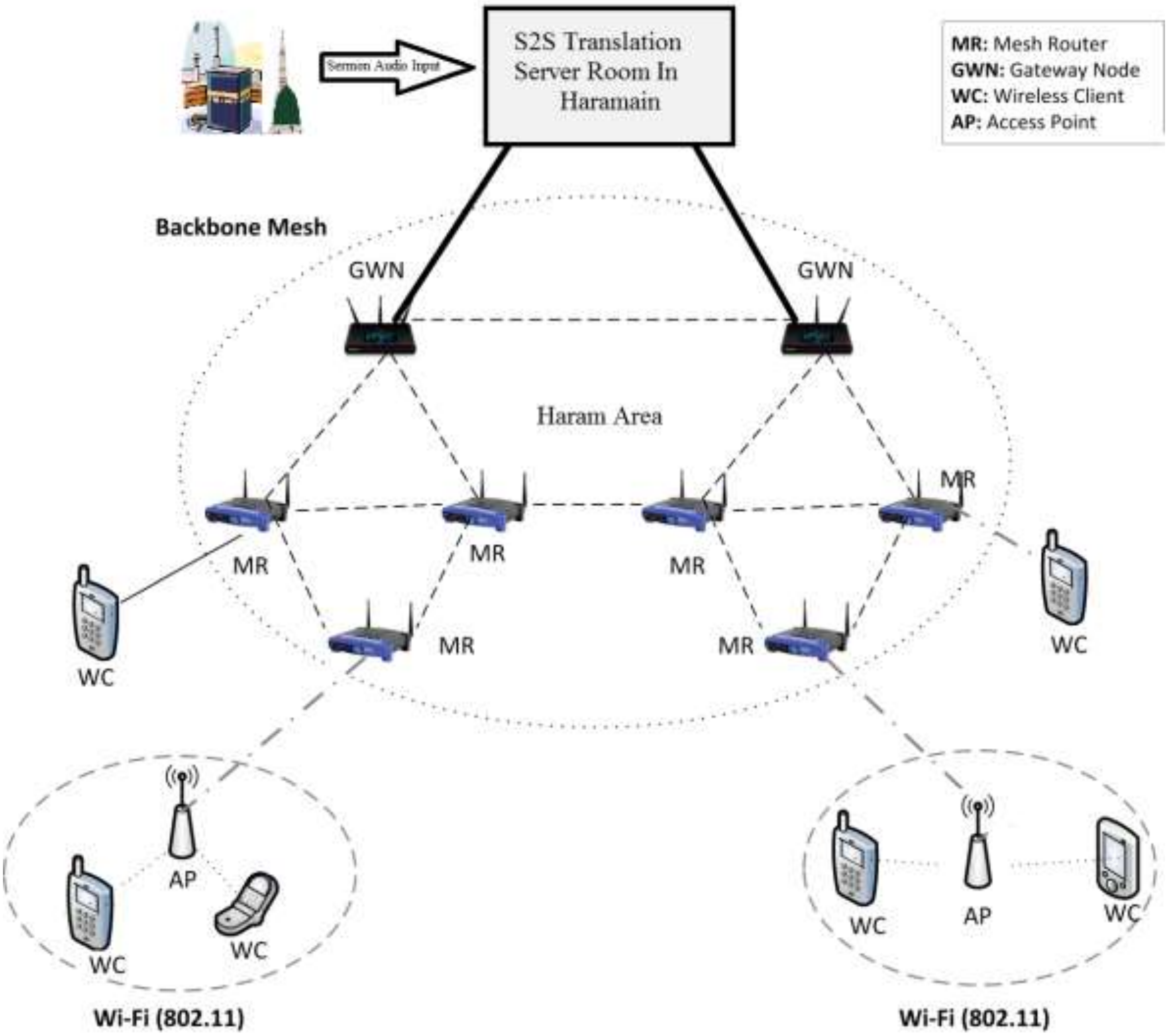

Figure 6. Architectural Overview for the Access to S2S Translation

Figure 5 explains the simplistic approach for the access to the translated Arabic speech in Urdu via IEEE 802.11 (WiFi) protocol. MR represents a mesh router, GWN represents the gateway node, AP access point and WC for wireless client. The key advantage of using this architecture is that without worrying about immense computational load on handhelds and mobile phones, the users can listen to the translated version of the Sermon (khutba) via preinstalled application on their phones, handheld PDA's. This architecture allows using the matured algorithms and approaches for SR, MT and TTS to be implemented on the Machines in Harmayn Server Rooms. The mobile users who wish to listen the Sermon (khutba) in real-time translated Urdu can receive the translated Sermon (khutba) on their mobile phones, handhelds and PDA's using the IEEE 802.11 protocol for WiFi. A WiFi Mesh Topology is proposed to accommodate the increased number of users present in the Haram Area. The high density areas can be find out by using RF tools such as Air Magnet planner etc. Use of Dual band AP's, overlapped AP design, load balanced traffic, Tuning of AP's power to avoid co-channel and adjacent channel interfacing and stress testing at maximum load can be useful designing strategies for such a dense network (NETGEAR, 2014).

6. CONCLUSION

A S2S translation for Arabic – Urdu, real-time translation mechanism is being proposed with in depth investigation for the appropriate corpora, ASR, MT and TTS for the required system. Moreover, meshed WiFi (IEEE802.11) is being proposed for the access technique for the users/listeners to listen to real-time translated sermons, the technique allows the matured protocols to be implemented without putting computational load on the handhelds, mobile phones and PDA's.

7. FUTURE WORK

Implementation of the proposed ASR, enriched corpora for both Arabic & Urdu which is easily available for both languages for the desired application i.e. Sermon translation which mainly consists Quranic and Hadith Arabic as well as Quranic and Hadith Urdu which is resource rich.

Implementation of SMT on the MT needs to be done whilst all the required components has been referred. TTS implementation can be done using the Architecture proposed. Lastly many improvements can be done for more robustness and increased capacity in the meshed WiFi based access technique which is proposed for the access for the users/listeners.

References

Actual Inventory of Pilgrims Hajj 1435, Central Department of Statistics & Information URL: http://www.cdsi.gov.sa/english/index.php?option=com_docman&Itemid=176

Adam Lopez, (2008). "Statistical Machine Translation," ACM Computing Surveys (CSUR), Volume 40 Issue 3, Article No. 8.

Bowen Zhou, Xiaodong Cui, Songfang Huang, Martin Cmejrek, Wei Zhang, Jian Xue, Jia Cui, Bing Xiang, Gregg Daggett, Upendra Chaudhari, Sameer Maskey, Etienne Marcheret, (2013). "The IBM speech-to-speech translation system for Smartphone: Improvements for resource-constrained tasks," Computer Speech and Language 27, 592–618.

H. Satori, M. Harti and Chenfour, (2007). "Introduction to Arabic Speech Recognition usin CMU Sphinx System," Information and Communication Technologies International Symposium proceeding ICTIS07 Fes

Hung Le Hong, Jackie Fenn, and Rand Leeb-du Toit, (2014.). "Hype cycle for emerging technologies 2014," Gartner Inc.

M. Alghamdi, A. H. Alhamid, and M. M. Aldasuqi, (2003). "Database of Arabic Sounds: Sentences," Technical Report, King Abdulaziz City of Science and Technology, Saudi Arabia, (In Arabic).

Mohammad A. M. Abushariah, Raja N. Ainon, Roziati Zainuddin, Moustafa Elshafei and Othman O. Khalifa, (2010). "Natural Speaker-Independent Arabic Speech Recognition System Based on Hidden Markov Models Using Sphinx Tools," International Conference on Computer and Communication Engineering (ICCCE 2010), Kuala Lumpur, Malaysia.

M. Ali, M. Elshafei, M. Alghamdi, H. Almuhtaseb, and A. Al-Najjar, (2008). "Generation of Arabic phonetic Dictionaries for Speech Recognition," IEEE Proceedings of the International Conference on Innovations in Information Technology, UAE, pp. 59 – 63,

M. Ali, M. Elshafei, M. Alghamdi, H. Almuhtaseb, and A. Al-Najjar, "Generation of Arabic Phonetic Dictionaries for Speech Recognition, (2008)." IEEE Proceedings of the International Conference on Innovations in Information Technology, UAE, pp. 59 – 63, NETGEAR Inc, "Best Practices for High DensityWireless Network Design In Education and Small/Medium Businesses," Available Online:

Rohit Prasad , Prem Natarajan, David Stallard, Shirin Saleem, Shankar Ananthakrishnan, Stavros Tsakalidis, Chia-lin Kao, Fred Choi, Ralf Meermeier, Mark Rawls, Jacob Devlin, Kriste Krstovski, Aaron Challenner, (2013). "BBN TransTalk: Robust multilingual two-way speech-to-speech translation for mobile platforms," Computer Speech and Language 27 475–491.

Sangmi Shin, Eric T. Matson and Jinok Park, Bowon Yang, Juhee Lee, Jin-Woo Jung, (2015). "Speech-to-Speech Translation Humanoid Robot in Doctor's Office," in Proceedings of the 6th International Conference on Automation, Robotics and Applications, Feb 17-19, Queenstown, New Zealand.

Sarmad Hussain, (2004). "Letter-to-sound conversion for Urdu text-to-speech system," Proceedings of the workshop on computational Approaches to Arabic Script based languages, Stroudsburg, PA, USA pg 74 – 79.

Quran corpus. Retrieved from http://corpus.quran.com/

Hadith Corpus. Retrieved from http://sunnah.com/

Quranic Urdu. Retrieved from http://www.quranurdu.com/

Archived Sermons of haramain, Retrieved from http://www.haramain.info/