*National Conference on Advances in Computer Science and Applications with International Journal of Computer Applications (NCACSA 2012)*
*Proceedings published in International Journal of Computer Applications® (IJCA)*

# TRANSLINGUATOR: Web based Application for Speech Translation of Human Voices Based on Voice Forensics without Changing the Source Voice

B.Naveen Raj,
Dept of ECE,
SMK Fomra Institute of Technology,
Kelambakkam-603 103, India.

T.Pearson, Ph.D.,
Professor,
SMK Fomra Institute of Technology,
Kelambakkam-603 103, India.

## ABSTRACT
Translinguator is a web based application that can be employed with the help of cloud computing technology. The application can be integrated with devices such as Mobile phones, tablets etc, or it can be designed with dedicated hardware as an independent device. It mainly involves the integration of various existing concepts in a specific sequence to obtain the unique desired output. This application is targeted in receiving the voice input of a speaker in a foreign language and translating it to the language known by the listener. The uniqueness of this application is that it can produce the translated output in a language which is known and desired by the listener (the person who is using the device) and that too in the same voice as the speaker using voice forensic analysis of the input speech. It employs the very common algorithms used in Speech recognition Engines (SRE), Text To Speech converter (TTS), Voice comparator & morpher, etc. The device/application can benefit a wide range of users such as students, travellers, PRO's etc. It is very economical as well as affordable to employ both as a dedicated device or an embedded web-based application.

## Keywords
cloud computing, speech synthesis, voice forensics, TTS (Text-to-Speech), web-technology, SRE (Speech Recognition Engine).

## 1. INTRODUCTION
Translinguator is a web based application for speech translation that comes into play in the day to day life necessities for a wide range of people. It has an interface that has a completely automated and web integrated technology [1] for translating speech from one language to another in the same voice itself with the help of voice forensics [2]. It acts very much like a human translator who helps us with translating an understanding what the other person says, except that it does not have limitations for the number of languages known. It is a one-time install and still more efficient as it can keep updating itself from the internet.

## 2. EXISTING SYSTEM & NEED FOR TRANSLINGUATOR
In today's scenario, there is no successful existing system for translation. Some models use the inbuilt database look-ups for translations. The really existing scenario is the hiring of translators who are well-versed in the required languages. The same system is being used in the parliaments of various countries, say, India, European countries and many more countries all over the world where people speak many languages [3]. Even the mighty UN assembly provides "Translation & interpretation services" with the help of an established organisation, consisting of people with certified multi-lingual skills.

These certified translators listen to what is spoken by the other person and translates/interprets them to their client through a microphone. The client is most probably a minister or a national leader listening to the interpretation. In such scenarios, inappropriate translation may lead to international issues & it is also a deciding factor of national security.

According to a survey, the maximum number of languages known by Guinness record holder named, Dr.Carlos do Amaral Freir is 115, when the total number of registered spoken languages is nearly 6500. So a speech translator becomes very necessary especially for people who travel the globe a lot.

In order to overcome this prevailing situation, we present you, TRANSLINGUATOR. The translinguator is proposed as a web-based application which will intensify its lingual knowledge i.e., the vocabulary database, with the help of neural networks enriched by the INTERNET. It will bring about a revolution in the areas where lingual constraints are a great problem. It can find its applications from schools & business & stretch its arms even to issues relating to national security.

## 3. SPECIAL CHARACTERISTICS OF TRANSLINGUATOR
1. *Unlimited lingual support* - This application can support practically unlimited languages as it derives all its necessary data from the WWW. It can support almost all the languages available over the internet & with the help of cloud computing technology, the unlimited support can be achieved by storing all the database information in the cloud server.
2. *Voice recognition* – It can also provide voice recognition features for security and also for forensics. It can be seen as a derived feature of the application.
3. *Artificial intelligence features* – The application also features artificial intelligence. It is strengthened by artificial neural networks in the vital blocks of the system [4] [5] and can auto detect the language of the speaker with few limitations.
4. *Complete web integration* – The application is completely integrated with the internet cloud for all its functions, except for input acquisition by the microphone input and output via speaker.
5. *No additional resources* – The application does not require any additional resources for its implementation. When it is embedded in a device like a mobile phone or

PC, these devices have built-in microphone & speakers. Hence no other additional resources will be specially required by the application.

# 4. TECHNOLOGY IN WORKING OF TRANSLINGUATOR

The translinguator, as an application, consists of 5 main blocks in its functioning. They are:

1. Speech acquisition
2. SRE (speech recognition engine)
3. Data uploader/ downloader
4. Online text translator
5. Speech synthesizer (TTS – Text To Speech converter)
6. Voice Morpher
7. Output of processed speech

In this paper, we shall discuss about the main attributes & functionalities of the above mentioned blocks.

## *A. Speech acquisition*

The first hardware element of the application is a microphone. The input to the application, i.e., the speech to be translated, is acquired by the microphone [6]. It is expected to have good noise cancellation feature, high SNR and directional array properties for better acquisition of the voice signal.

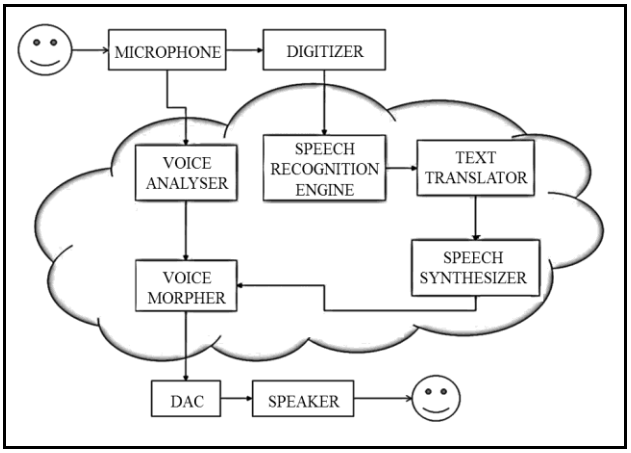


**Fig. 1 Block diagram of translinguator**

This input is digitized and processed. The processing in this level involves various speech processing techniques. One of the most important processing to be done in this point is track separation [7]. When more than one person's voice is received by the microphone or when the an external disturbance gets mixed with the human voice, the application is smart enough to split the tracks and performs further processing on the human voice track.

## *B. Data uploader*

This human voice track is alone uploaded to the internet cloud using some the data uploaders. The further processing of the signal is performed over the internet. This is done to avoid local processing overload and resource requirements.

## *C. Speech recognition Engine*

Once the data is uploaded to the cloud, the voice data is first given to the online speech recognition engine (SRE) [8]. A speech recognition engine is an application which converts speech to text. These online SRE's have the advantage that they are trained by a variety of accents and hence they have strong and enriched database. The Speech recognition engine performs three main functions. They are as follows:

- Recognition of the language in the isolated input voice (optional).
- Adaption to the recognition process based on the accent and attributes of the received voice [9].
- Conversion of the phonetics in the speech to text (of same language), which is mostly the phonemes of the particular word (often expressed in English).

The following diagram shows a simplified block diagram of the speech recognition process.

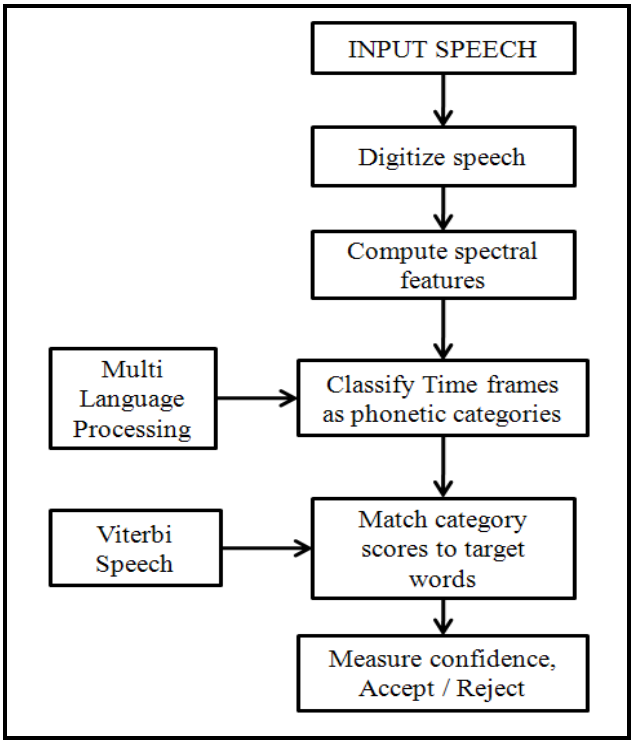


**Fig. 2 Block diagram of SRE**

The output of the SRE is textual form of the speech in that language. It is performed by computing the spectral features & analysing the phonetic categories based ion MLP(multi-language processing). The SRE is not a complex application and hence it can also be localized in the application-installed device itself to convert the speech to text and then it can be uploaded to the cloud.

## D. *Online text translator*

The next step is the translation of the speech recognized text to the desired language, which is set by the user. The user sets the language he knows/wishes, as the target language. This textual translation is performed by online translators like GOOGLE TRANSLATOR [10], or MICROSOFT BING TRANSLATOR, etc. They are open source applications available in the internet for textual translation across many languages [11]. Google translator supports 63 languages till date and it is still growing.

The application is self-aware & hence supports the "Auto-detect" feature to find the source language [12] [13]. The phonetics of the source text is sensed by the application and translated to the user-desired target language, automatically.

In this level, the text is translated from one language to another and passed on for further processing.

*E. Speech synthesizer*

The subsequent processing involves conversion of this translated speech to voice. This is done by a speech synthesizer [14]. It is an application which transforms text to speech, based on phonetic analysis. It performs the reverse process of a speech recognition engine. For example, Talk_It! [15] is a simple application which reads out the entered text based on the phonemes. The application splits the words into phonemes and reads them out. By this approach to words, a speech synthesizer is able to read the words from any language.

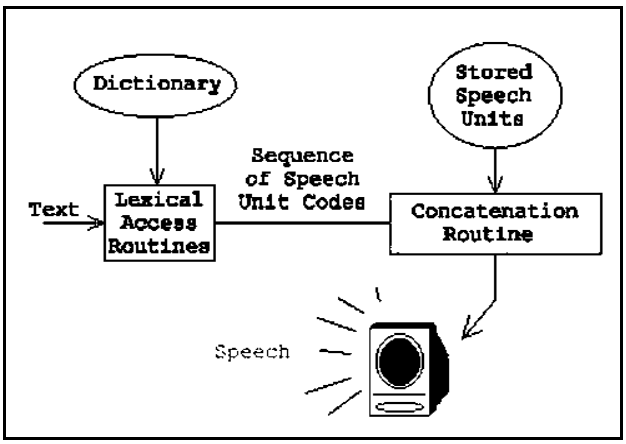


**Fig. 3 Block diagram of speech synthesizer**

*F. Voice morpher & final output to speaker*

The outputs of such speech synthesizers are robotic voices and hence they are monotonous. So it is often boring and unpleasant to hear. Hence, another block is added to the application – VOICE MORPHER [16].

The voice morpher is a very interesting application whose functioning is mainly concentrated in duplicating voices. This application extracts samples of phonetics from all the inputs and enhances the database with these phonemes. Training of voice morpher is also necessary just like speech recognition engines. It is done by neural networks. The translated and synthesised monotonous voice is now morphed with the help of samples of the original voice from the phonetics database.

The output of the voice-morpher is actually the input message, which is translated to the desired language, and is given as output through the speaker in the same original voice by forging the voice samples obtained by forensics. A few smoothening and filtering operations are performed on this output voice and then the final output is given to the speaker. This is the technology behind the working of the TRANSLINGUATOR.

## 5. PERFORMANCE ANALYSIS

On an average, a recording of about 30 seconds by a high definition microphone of normal specifications requires a storage space of about 512kb. This data has to be uploaded to the internet for the translinguation process. The following tabulation shows the performance analysis of the application based on the response time across various connection types with various data rates.

TABLE I. **PERFORMANCE ANALYSIS**

| Connection type | Data rate | Application response time (in seconds) |
|---|---|---|
| Dial-up | 4 kbps – 128 kbps | 600 – 250 |
| Broadband | 256 kbps – 2 Mbps | 100 – 50 |
| 3G | 4 Mbps – 20 Mbps | 50 – 20 |

The application response time changes drastically from dial-up to broadband connection but does not show much variation between broadband and 3G networks. This is because the application response times is the summation of upload time, online processing time (latency) and download time. The variation is the connection speeds can effect on the upload and download times but the online processing time is based on many unpredictable factors like internet traffic, the processing capability of the server, strength of the neutral network in the involved languages, etc. By improving the processing capabilities of the cloud server and using dedicated internet lines, the latency can be reduced very well to the scales of milliseconds.

## 6. LIMITATIONS

Despite the many advantageous features and characteristics of the translinguator, there are also few limitations. Some of them are as follows:

- The voice amplitude and accent of different people vary from person to person. So training the SRE is a time-consuming process.
- SRE's have not been developed for many languages except a very few popular languages.
- Online translators have some bugs in the translating processes and it tends to make mistakes in 'auto-detect language' feature.
- The improvisation of the processing time requires high processing capabilities & high data rates.
- The TTS (text-to-speech converter) has also not been developed for many languages and the phonetic synthesis is a difficult process.
- Sampling the voice samples using voice morph and creating a voice disguise is a tedious process.

## 7. CONCLUSION

The translinguator discussed in the paper is as an application. It is a well known that dedicated hardware is faster than software. It can be designed as a device which can satisfy all the requirements of the application. Some hardware specifications to be taken care are microphone, processor, memory, wireless network adapter, etc.The application is very user friendly & enriches its database from the internet with neural networks. It can have great advancements with the further developments and advancements of various technologies in the future.

# 8. REFERENCES

[1] Rajkumar Buyya, Chee Shin Yeo, and Srikumar Venugopal "Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities" The 10th IEEE International Conference on High Performance Computing and Communications, pp.10 Table 1

[2] Zahorian,S.A.; Livingston,D.; Dept. of Electr. & Comput. Eng., Old Dominion Univ., Norfolk, VA, "Neural networks for feature computations in automatic speech recognition" in Neural Networks, 1992, IJCNN., International Joint Conference.

[3] Richard Corbett, Francis Jacobs, Michael Shackleton in "The European Parliament" 7th edition, page 39 -41.

[4] Dupont,S.; Cheboub,L; TCTS-MULTITEL, Faculte Polytech. de Mons, "Fast speaker adaptation of artificial neural networks for automatic speech recognition" in Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference.

[5] Bodenhausen,U.; Hild,H.; Dept. of Comput. Sci., Karlsruhe Univ., "Automatic construction of neural networks for special purpose speech recognition systems" in Acoustics, Speech, and Signal Processing, 1995. ICASSP-95.

[6] Jacob Benesty, Jingdong Chen, Yiteng Huang in Microphone array signal processing, 2008 edition.

[7] Kusumoputro, B.; Triyanto, A.; Fanany, M.I.; Jatmiko, W.; Fac. of Comput. Sci., Univ. of Indonesia, "Speaker identification in noisy environment using bispectrum analysis and probabilistic neural network" in Computational Intelligence and Multimedia Applications, 2001. ICCIMA 2001.

[8] David Ferrucci, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A. Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John Prager, Nico Schlaefer and Chris Welty in Learning, Knowledge, Speech recognition, Text recognition, Lexical formulation, Translation (2010)

[9] Kung-Pu Li; ITT Aerosp. Commun. Div., San Diego, CA, "Automatic language identification using syllabic spectral features" in ICASSP-94.

[10] Google translator http://translate.google.com/support/

[11] Wenjie Li, Diego Mollá-Aliod, "Computer Processing of Oriental Languages: Language Technology for the Knowledge-based Economy" in ICCPOL '09, Hong Kong.

[12] Farris,D.; White,C.; Khudanpur,S.; Center for Language & Speech Process., Johns Hopkins Univ., Baltimore, MD, "Sample selection for automatic language identification" in ICASSP 2008.

[13] Tian, J.; Suontausta, J.; Speech & Audio Syst. Lab., Nokia Res. Center, Tampere, Finland, "Scalable neural network based language identification from written text" in ICASSP '03.

[14] Acero, A.; Speech Technol. Group, Microsoft Corp., Redmond, WA, "An overview of text-to-speech synthesis" in Speech Coding, 2000

[15] Turk, O.; Schroder, M.; Sensory, Inc., Portland, OR, USA, "Evaluation of Expressive Speech Synthesis With Voice Conversion and Copy Resynthesis Techniques" in Audio, Speech, and Language Processing, IEEE Transactions, July 2010.

[16] Hashimoto, K.; Yamagishi, J.; Byrne, W.; King, S.; Tokuda, K.; Dept. of Comput. Sci. & Eng., Nagoya Inst. of Technol., Nagoya, Japan, "An analysis of machine translation and speech synthesis in speech-to-speech translation system" in ICASSP '11.