

# Speech Signal Transformation with FS Variation Technique for Military Purposes

Azmaliza Azizi, Muhammad Suhaimi Sulong, Tasiransurini Ab Rahman, Nurulhuda Ismail and  
 Ariffuddin Joret

**Abstract**— Security in human identification system has become a major concern nowadays, especially for sending information verbally. Therefore, the need of protecting the internal factor especially human speech is a main objective in this study. The characteristics of the human speech using several methods has been continuously practiced are outlined, for instance cepstral analysis, linear prediction coding, and the Pitch Synchronous Overlap-Add (PSOLA) method. This paper described the used of frequency sample variation, FSV technique to transform speech signal to a deeper, higher or lower tone. The analysis using PSOLA method with this technique and Matlab as GUI has been performed. The developed system is capable of transforming the input speech into three types of output speech namely male, female and child. The results show that the system is successfully and effectively portray different people with realistic gender changes to conceal the identity. It is said to be a potential system for any kind of security particularly by disguising the character of the voice. The alterations with other voice variable are going to be a further development.

**Keywords:** Speech signal transformation, PSOLA method, Matlab

## I. INTRODUCTION

HUMAN identification system, nowadays, is a very vital for security purposes [1] and currently, it is a major concern in military purposes. The most important part is sending information verbally without expose the speaker identity. There are differences in human voice frequency ranges between male and female. The voiced speech of a typical adult male will have a fundamental frequency of

This work was supported in part by Universiti Tun Hussein Onn Malaysia (UTHM).

Azmaliza Azizi was with Universiti Tun Hussein Onn Malaysia. She is now with the Department of Research and Development, Agilent Technologies Sales (Malaysia) Sdn Bhd ... Penang Service Store. 17-G-6, Medan Kampung Relau. 1201 Piner Road. Bayan Point 11900 (corresponding author, phone: 013-7599319; e-mail: azmaliza\_azizi@yahoo.com.my).

Muhammad Suhaimi Sulong and Tasiransurini Ab Rahman are with the Department of Computer Engineering, Faculty of Electrical and Electronics Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Johor, Malaysia (e-mail: msuhaimi; surini@uthm.edu.my).

Nurulhuda Ismail and Ariffuddin Joret are with the Department of Communication Engineering, Faculty of Electrical and Electronics Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Johor, Malaysia (e-mail: nhuda; ariff@uthm.edu.my).

from 85 to 155 Hz, and that of a typical adult female from 165 to 255 Hz. The female voice is actually more complex than male due to differences in the size of the vocal cords and larynx between male and female and also due to greater natural melody in their voices. This causes a more complex range of sound frequencies than in a male voice. [2,3]

Many researchers in the area of digital signal processing have focussed on speech processing [1] especially in the topic of speech signal transformation, or in other word, voice conversion [1,4,5]. The aim is to transform the speech of a source speaker to a target speaker. Other applications also have interest in this voice transformation such as foreign language training, movie dubbing [1,6], broadcasting, Karaoke [7], voice editing for films, speech synthesis, speaker verification [4,5], language study [8], multimedia and music [9].

According to D. Rentzos *et al.*, there are three essential components in order to have effective voice conversion system, namely voice feature extraction, model estimation and voice mapping [4,5]. In addition, vocal tract transfer function and pitch range are the characteristics of human voice which are different for different sound. The methods used to transform speech signal with vocal tract transfer function are cepstral analysis and linear prediction coding whereas PSOLA method used to change the pitch voice [1].

## Cepstral Analysis

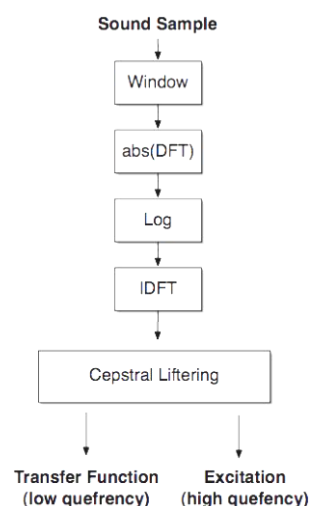


Figure 1: Cepstrum block diagram

The use and application in this analysis is “cepstrum” (spectrum) domain. It is a common transform to perform as “cepstral” (spectral) analyzer to gain information from one person’s speech signal. Within human speech, the voiced (transfer function) and unvoiced (excitation signal) sounds are employed to form words. After having calculated the cepstrum with these words, the signal will have to “lifter” (filter), separating the speech signal components; transfer function (which contains the vocal quality) and excitation (which contains the pitch and the sound) signal. The cepstrum block diagram is shown in Figure 1. It is the full process used to compute the cepstrum. [10,11]

*Linear Prediction Coding (LPC)*

A method that predicts a sample of a speech signal based on several previous samples. It is similar with the cepstrum method, where it uses LPC to separate the transfer function and the excitation. In the time domain, both components of speech signal are convolved to create the output voice signal. [11-13] Figure 2 shows the LPC algorithm where the original signal that put through the filter (transfer function) to get the excitation component. In order to get back the original signal just simply putting the excitation component through the inverse filter. [12]

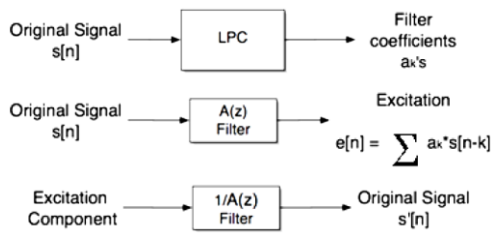


Figure 2: LPC algorithm

*PSOLA*

This method modifies the pitch of a speech signal but it maintaining the other vocal qualities. Figure 3 shows the three steps of basic algorithm for the PSOLA technique. The result is a signal with the same spectrum as the original but with a different fundamental frequency. [11,14]

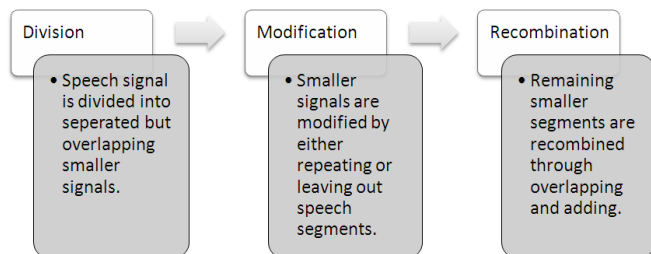


Figure 3: PSOLA technique

II. PROJECT DEVELOPMENT

The speech signal transformation system is developed using Matlab as the GUI and PSOLA method with FSV

technique. The block diagram of the developed system is illustrates in Figure 4.

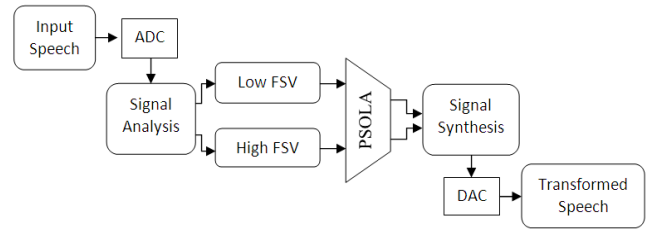


Figure 4: Block diagram of speech signal transformation

In this system of speech signal transformation, the main alteration is the pitch on an input speech. The speech transformation is properly performed depends on the frequency characteristics of input speeches. The pitch will be shifted toward a lower or higher frequency on the frequency axis.

The default value of frequency sample rate used is 11025Hz. When the input speech frequency is adjusted, it will affected the time of the output speech as it proven that the frequency is inversely proportional to the time. Therefore, the output speech will sound faster or slower depends on FS varied, hence transformed the input speech. As a result, the transformed speech will probably produce an unnatural human voice if the transformation is not properly performed.

III. RESULTS AND DISCUSSIONS

The developed system is capable of transforming the input speech into three types of output speech; namely male, female and child as shown in Figure 5. The input speech was transformed to a required output speech by first selecting the types of output speech.

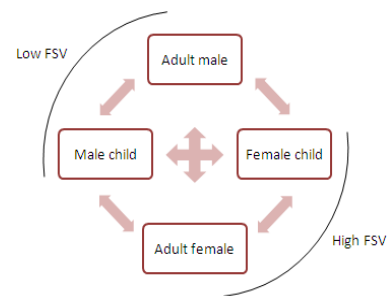


Figure 5: Input and transformed speech signal results chart

As a sample of input speech signal, adult female voice was used to transform the speech to female child and adult male as shown in Figure 6 and 7 respectively. The input speech sound sample was “Hello”. The transformed speech of a female child had higher pitch and faster rate than the input speech, conversely, lower pitch and slower rate went to the adult male transformed speech.

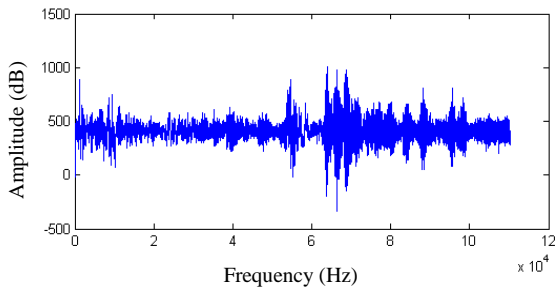


Figure 6: Transformed signal adult female to female child

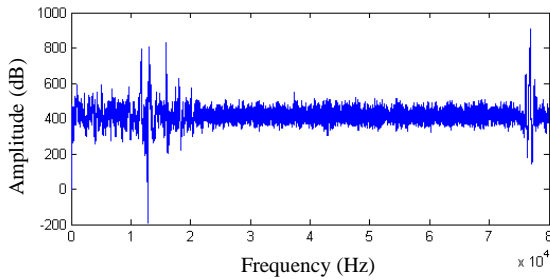


Figure 7: Transformed signal adult female to adult male

#### IV. CONCLUSION

Human voice has different sound and tone between age and genders and it is important to conceal the identity of the speaker by disguising the character of the voice for security and military purposes. A system aim to transformed input speech to a specific desired speech has been developed using PSOLA method with FSV technique. The results show that the system is successfully and effectively portray different people with realistic gender changes; however, the system had only predetermined the output speech with fixed FS value. Further development will also focus on having variable of FS in order to comprise various output speech.

#### ACKNOWLEDGMENT

The authors would like to thank the UTHM for their encouragement and acknowledge the financial support received from the government of Malaysia.

#### REFERENCES

- [1] Gina Upperman. (2004, December). Introduction to methods for voice conversion. Connexions module, Rice University, Houston, Texas. Available: <http://cnx.org/content/m12479/1.4/>
- [2] I. R. Titze, *Principles of Voice Production*. USA: University of Utah, 2000.
- [3] R. J. Baken, *Clinical Measurement of Speech and Voice*. Singular Publishing Group, 1996.
- [4] D. Rentzoa, S. Vaseghi, E. Turajlic, Q. Yan, C-H. Ho, "Transformation of speaker characteristics for voice conversion," in *2003 IEEE ASRU*, pp. 706–711.
- [5] D. Rentzoa, S. Vaseghi, Q. Yan, C-H. Ho, "Voice conversion through transformation of spectral and innotation features," in *2004 IEEE ICASSP*, pp. I-21–I-24.
- [6] O. Turk, L. M. Arslan, "Subband Based Voice Conversion," in *Proc. ICSLP*, Denver, Colorado, 2002, pp. 289–292.
- [7] W. Verhelst, T. Ceyskens, P. Wambacq, "On inter-signal transplantation of voice characteristics," in *Proc. 3<sup>rd</sup> IEEE Benelux Signal Processing Symposium*, Belgium, 2002, pp. 137–140.
- [8] Y. H. Jung, M. K. Ham, C. J. Jung, M. J. Bae, "PSOLA pitch alteration of accuracy multiple using voice/unvoice speech detection algorithm," in *2002 Proc. IEEE TENCON Conf.*, pp. 901–904.
- [9] Oytun Türk, "New methods for voice conversion," M.Sc. thesis, Dept. Electrical and Electronics Eng., Boğaziçi Univ., Istanbul, Turki, 2003.
- [10] Brian Van Osdol. (2004, December). Cepstrum. Connexions module, Rice University, Houston, Texas. Available: <http://cnx.org/content/m12469/1.4/>
- [11] Gold, Ben and Nelson Morgan, "Speech and audio signal processing: Processing and perception of speech and music," John Wiley and Sons, Inc: New York, 2000.
- [12] Gina Upperman. (2004, December). Linear prediction coding in voice conversion. Connexions module, Rice University, Houston, Texas. Available: <http://cnx.org/content/m12473/1.4/>
- [13] Deng, Li and Douglas O'Shaughnessy, "Speech processing: A dynamic and optimization-oriented approach," Marcel Dekker, Inc: New York, 2003.
- [14] Gina Upperman. (2004, December). Changing pitch with PSOLA for voice conversion. Connexions module, Rice University, Houston, Texas. Available: <http://cnx.org/content/m12474/1.4/>