



# Speech Transformation

**Mr. Mohit Kumar**

M.E. Scholar (Comm.) SSGI, Bhilai(C.G.)

**Mr. Piyush Lotia**

Associate Professor, HOD (E&I) SSGI, Bhilai(C.G.)

ABSTRACT

Speech transformation is a process of transforming the characteristic of speech uttered by a one speaker i.e. source speaker, such that a listener would believe that the speech was uttered by another speaker i.e. target speaker. We can also say that it is a process of modifying the speech signal of one speaker (source speaker) so that if it had been pronounced by a different speaker(target speaker).Over last few years, the interest in speech transformation has risen significantly due to its wide application. Two main aspects of the transformation problem are speech quality and intonation.

The speech quality transformation component of the system has two main parts. The first component transform the spectral envelop and this transformation is done by Gaussian mixture model which is trained on aligned speech from source and target speakers. The second part of the system predicts the spectral details from the transformed LPC parameters. A novel approach is proposed which is based on a classifier and residual codebooks. A listening test shows that the proposed method makes it possible to convert speaker's individuality while maintaining high quality.

**KEYWORDS**

speech transformation; source speaker; target speaker; LPC parameters.

**I. Introduction**

The aim of speech transformation is to transform the sound of source speaker as sound of target speaker. Speech signal has a wide range of information. Speech modification techniques attempts to transform the speech signal uttered by a given speaker so as to alter the characteristics of his or her speech. As the psychoacoustic correlates of speaker identity remain largely unknown, it is often necessary to specify the desire modification of speech characteristics with reference to the target speaker. So the modification of speech of one speaker so that it sounds as if it was uttered by another speaker is generally known as speech transformation.

speech transformation aims at the control of non-linguistic information of speech signals such as speech quality and speech individuality. It refers to the various modifications one may apply to the sound produced by a person. speech transformation covers a wide area of research from speech production modeling and understanding to perception of speech, from natural language processing, modeling and control of speaking style to pattern recognition and statistical signal processing.

Speech transformation is performed in two steps. in training stage, acoustic parameters of speech signals uttered by both source and target speakers are computed and appropriate rules mapping the acoustic space the source speaker on to that target speaker are obtain. In the transformation stage, the acoustic features of the source signal are transformed using the mapping rules such that the synthesized speech processes the personalities of the target speaker.

**II. Application**

In this literature no. of application for speech transformation are given which are.

- The most important application of speech transformation is the use as a module in text to speech synthesis, where we want to change the standard speaker's speech characteristic so that we suppose to hear another speaker, target speaker.
- In embedded environment, speech transformation can serve to manipulate speech in such a way that the original speaker can not be recognized or the speaker's gender or age is changed. Due to computational limitations. Conversion algorithm of embedded system focus on manipulating only a few parameters as fundamental frequency and vocal

truck length.

- Cross language speech transformation has been applied to dubbing task in the film industry .In the next future it is to be used inside a speech to speech translation framework, where a translated sentence is uttered with the source speaker.
- Speech transformation is also used in personalized human computer interface where generating a virtual copy of speech that are meaningful to the user e.g. a mother's speech in educational software.

Speech transformation is widely used in entertainment to generate special effect and also used in gaming to generate different speech.

**III. Approaches**

There is vast number of approaches proposed till now. In this paper only the most recognized once are taken into account.

- The inventors of speech conversion Abe et al, clustered feature vectors of source and target speaker utterance and tried to find an appropriate mapping between this classes. This vector quantization approach has been extended by Arslan and is still used in state of the art speech transformation.
- The most popular technique is the application of the linear transformations to the spectra of pitch synchronous frames that are coded using mel frequency cepstral coefficients or line spectral frequencies.
- The paper by Zheng et al suggested a new data alignment method for text independent speech conversion considering both the phonetic accuracy and preservation of the internal topology of parameter space. To insure accuracy they use phonetic labels of the training data as supervisory information and phonetic restriction is considered for supervised alignment.
- The paper by David Sundermann suggested text-independent and cross language speech conversion in which parallel database are available for training purposes. Main problem in this conversion is that both the source and target speaker speaks different languages.
- Over the last few years, remarkable efforts were taken to describe the speech transformation applying formant and anti-formant parameters.

#### IV. Conclusion

Speech transformation covers a wide area of research from speech production modeling and understanding to perception of speech, from natural language processing, modeling and control of speaking style, to pattern recognition and statistical signal processing. To improve further the quality of speech transformation systems, more effort should be made to take into account the nonlinear phenomena during the speech production process and result from the natural language processing area. More accurate, flexible and meaningful for speech, model should be developed for modeling longer speech segments than current models do. However, it is evident that speech transformation requires more than just modeling the speech signal; it requires understanding the speech process in terms of production, perception and natural language processing.

#### V. Future Work

From last few years work on speech transformation becomes a difficult task because of its some limitation. In speech transformation we need to map the characteristics of both source and target speakers and the characteristics of source speaker can not be constant it may vary time to time. For example the frequency of our speech for same set of sentences may be different and amplitude of our speech may also be vary.

Many researchers report that speech converted speech has a muffling effect possibly because of the broadening of spectral peaks in the converted speech and lost of spectral details in the converted spectra. To cope with this effect, statistical approaches using the global variance of converted spectra in each utterance have been proposed and estimation of the lost spectral details.

Speech transformation is a very new field of research and many works have done in this field for many applications given above. In future speech transformation will be used in numerous and versatile applications after addressing the challenges given in above section.

#### REFERENCES

- [1] Yannis Stylianou, "Voice transformation", Institute of computer science, FORTH, and Multimedia information lab, CSD, UoC, Greece, .IEEE, 2009. | [2] Alexander Kain and Michael W. Macon. "Design and evaluation of a Voice conversion algorithm based on spectral envelope mapping and residual prediction", Center for Spoken Language Understanding (CSLU) 20000 NW Walker Road, Beaverton, 97006, USA. | [3] Alexander Kain, Akiko Kusumoto, John-Paul Hosom, Jan van santen, "Voice Transformation, Increasing speech Intelligibility for the Hearing and Speaking impaired", Center For Spoken Language Understanding Oregon Health and Science University, May 2010. | [4] Ki Seung Lee, Dae Hee Youn , Il Whan Cha, "A New Voice Transformation Method Based On Both Linear and Nonlinear Prediction Analysis", Department of Electronic Eng., Yonsei University Seoul, 120-749, Korea. | [5] Zdenek Hanzlicek and Jindrich Matousek. "Voice conversion based on probabilistic parameter transformation and extended inter-speaker residual prediction", University of West Bohemia, Faculty of Applied Science, Dpt. of Cybernetics, Univerzitni 8, 306 14 Plzen, Czech Republic. | [6] David Sundermann, "Speech conversion: State-of-Art and future work", Universitat Politècnica de Catalunya, 08034 Barcelona, Spain. | [7] H. Kuwabara and Y. Sagisaka. "Acoustic characteristics of speaker individuality: Control and conversion", Speech Communication, 16(2):165-173, 1995. | [8] Yannis Stylianou, Tomoki Toda and Alexander Kain, "Introduction to the special section on Voice Transformation", IEEE Transactions and audio, speech, and language processing, vol. 18. No.5, July 2010.