To Investigate the Relation between Class Dependent Transformation Function Coefficients for Voice Conversion

Radhika Khanna, Parveen Lehana

Abstract— Voice conversion, is the modification of the speech of one speaker (called source speaker) into the speech of another speaker (called target speaker). The quality of the synthesised speech depends upon the precise estimation of the transformation function, which is very difficult as there are many features of speech which are difficult to extract automatically, such as meaning of the passage and intention of the speaker. Although a good estimate of the transformation function may be obtained from the dynamics of the spectral envelopes of source and target speakers , the misalignment of the patterns of phonemes in the passages of the source and target may hamper the precise estimation. The quality of the transformed speech depends upon various factors such as alignment of phonemes of source and target speaker, estimation of transition segments, estimation of transformation function, and resynthesis algorithm, the quality of the speech will be further deteriorated if the transition segments are not aligned properly. The most error prone process in voice conversion is the alignment of transition segments. It means exact alignment of the corresponding speech units in the source and target passages is necessary for the accurate estimation of the transformation function. Different methods have been used for alignment. These may be based on dynamic warping, cluster mapping, HMM, speech recognition, unit selection, and adaptation models.

Index Terms— DTW, HMM, GMM, RBFN, ANN, VQ, MLM, LPC, MFCC

1 INTRODUCTION

The area of research dealing speech signals may be broadly divided into three sub fields: speech recognition, speech synthesis, and speech processing. The hot area of speech processing, i.e. voice conversion, is the modification of the speech of one speaker (called source speaker) into the speech of another speaker (called target speaker) [1]-[4]. Voice conversion is used in many applications namely dubbing, to enhance the quality of the speech, text-to-speech synthesizers, online games, multimedia, music, cross-language speaker conversion, restoration of old audio tapes, cellular applications, low bit-rate speech coding, etc. Voice conversion is carried out using a speech analysis-synthesis system, in which the parameters of the source speech are modified by a transformation function and resynthesis is carried out using the modified parameters. The transformation function is obtained by analyzing the aligned source and target speakers' utterances. Various techniques are used for voice conversion such as codebook based transformation [2], [3], dynamic frequency warping technique [5]-[7], speaker interpolation [8], artificial neural networks (ANN) [9], Gaussian mixture models (GMM) [10]-[12], hidden Markov model (HMM) [13], and vector quantization (VQ) [14].

There are two main stages in voice conversion: training and transformation. The training stage may further be divided into three steps: acoustic modeling, segmentation and alignment, and acoustic mapping. In the acoustic modeling stage, speaker-specific parameters are extracted from the speech waveform. These parameters describe the short-term and long-term characteristics of the source and target voices. Vocal tract, glottal source (pitch, spectral tilt, open / closed quotient), duration, and energy characteristics convey important speaker-specific information [15]-[19]. Linear prediction coefficients (LPC) [20], line spectral frequencies (LSF) [21], mel-frequency cepstral coefficients (MFCC) [22], formant frequencies and bandwidths [23], and sinusoidal transform coding (STC) parameters [24] may be used for modeling the vocal tract characteristics. There has been considerable amount of work on the analysis, modeling and modification of glottal source characteristics in voice quality research [25]-[27]. Pitch is one of the most important speaker-specific dimensions among the glottal source characteristics. It may be estimated using the autocorrelation function, average magnitude difference function, Fourier transform, and harmonic analysis [28]. Dynamic programming is a commonly employed method to avoid discontinuities and hence improve the robustness of the pitch detection algorithm [29].

The second step, alignment, is necessary to determine corresponding units in the source and target voices. This is due to the fact that the durations of sound units (i.e. phonemes or sub-phonemes) may be quite different among speakers. It is preferable to employ automatic alignment techniques like dynamic time warping (DTW) [30] and HMM [31] because manual alignment is time consuming.

The final training step is the estimation of the acoustic mapping function between the source and the target speaker's acoustic spaces using machine learning techniques like vector clustering/quantization [14], codebook mapping [32], weighted codebook mapping [3], [33], GMM[11], radial basis function networks (RBFN) [34], ANN [9], and self-organizing maps (SOM) [35]. The main distinction between the earlier methods [14], [32] and more recent methods [3], [11], [33] is that smoothing among the mapping units is performed to reduce distortion at frame boundaries. Another difference of more recent methods is the employment of text and language independent automatic techniques for alignment such as sentence-HMM and DTW.

The transformation stage employs acoustic analysis techniques similar to the acoustic modeling step in training.

Once the parameters of the input waveform are estimated, speaker transformation rules are imposed to obtain the corresponding target parameters. Necessary modifications are performed on the input waveform to match the target speaker characteristics. The modifications include transformation of the vocal tract, glottal source, duration, and energy characteristics. The vocal tract characteristics can be transformed using formant modification [36], interpolation of the line spectral frequencies [3], and sinusoidal modeling techniques [37].

The objective of this paper is to investigate the relation between class dependent transformation function coefficients. Investigation was carried out to study the similarity of the transformation functions for different phonemes.

2 Methodology

The quality of the transformed speech depends on the accuracy of transformation function. To obtain the accurate transformation function the speech signals of source and target speakers must be aligned properly. Fig. 1 shows the block diagram of the methodology used. Methodoloy is divided into following three phases:

2.1 Speech material

Speech data is required for both training and testing. Speech material was recorded from eight speakers (4 male and 4 female, ages: 20-23years). The male speakers are referred to as M1, M2, M3, and M4 and the females F1, F2, F3, and F4. The speakers in our experiment were university students of the same age group and had Hindi as their first language. It is desirable that the speakers belong to same group in terms of language to avoid accent related bias. The material was recorded in an acoustically treated room with 16 kHz sampling and 16-bit quantization rate. The recorded material was manually segmented into sentences, and 50 sentences which were considered to be correctly articulated by all the speakers were selected for use as the speech utterances for voice conversion experiments.

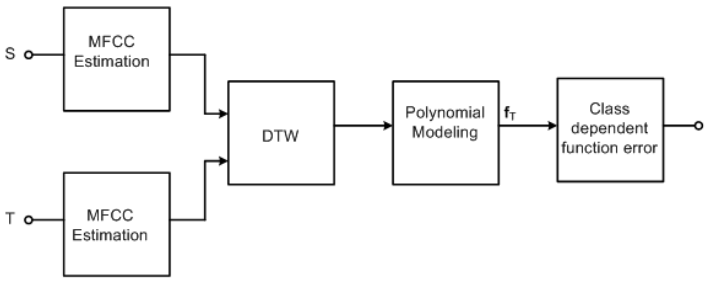

Fig. 1 Block diagram of the Methodology

2.2 Estimation of transformation function using DTW

The utterances were converted to mel frequency cepstral coefficients (MFCC) based feature vectors. The feature vectors were aligned using DTW. Each parameter for generating the target speech is modelled as a multivariate polynomial func tion of all the parameters of the source speech, and the sets of these polynomial functions is obtained by analyzing a set of time aligned source and target frames. MLM based transformation functions were estimated.

.

2.3 Estimation of class dependent function error

To investigate the similarity of the transformation functions for different phonemes, they are divided into 8 classes usin VQ and then class dependent function errors are estimated.

3 Results

To investigate the similarity of the transformation functions for different phonemes, they are divided into 8 classes. Investigation was carried out with four speaker pairs (F1-F2, F3-M3, M4-F4, and M1-M2).

Table 1 (a) show the component functions (F1-F10) of the mapping for male to male (M1-M2) transformation. The first component function (F1) has been derived from the given training data using the MFCC coefficient of the source speaker and only first MFCC coefficient of the target speaker.

For function F1, the values of the function coefficients representing the function $F_1 = c_{0,1} + c_{1,1}x_1 + c_{2,1}x_2 + \ldots\ldots + c_{n,1}x_{p,}$ comes out to be same for the phonemes

i. क and च
ii. ट, ड, फ, स, श, and ष
iii. ऊ, औ, घ, ण, छ, ठ, ढ, ध, न, म, र, ल, व, and ह
iv. अ, आ, इ, ई, ओ, ए, ऐ, ख, त, थ, द, ब, भ, य, and क्ष

Phonemes ज, ग, प and झ do not show any similarity w.r.t. to function coefficients with any other phonemes.

For function F2, the values of the function coefficients representing the function $F_2 = c_{0,2} + c_{1,2}x_1 + c_{2,2}x_2 + \ldots\ldots + c_{n,2}x_{p,}$ comes out to be same for the phonemes

i. झ and ठ
ii. उ, द, and ल
iii. ड, प, भ, and म
iv. ई, ओ, च, ज, and य
v. क and ग
vi. इ, ए, छ, थ, ब, श, and ष
vii. अ, आ, ऊ, ऐ, ओ, ण, and न
viii. घ, ट, ढ, त, ध, फ, र, स, and क्ष

• , PH-91-9419201579. E-mail: radhikakhannaguptae@yahoo.in
• Co-Author PH- 91-941912000267, E-mail: parveenlehana@gmail.com

For function F3, the values of the function coefficients representing the function ($F_3 = c_{0,3} + c_{1,3}x_1 + c_{2,3}x_2 + ... + c_{n,3}x_{p,}$) comes out to be same for the phonemes

i. क and ग
ii. इ, ई, ओ, औ, ण, च, ध, ज, र, and ह
iii. अ, आ, उ, ए, घ, ज, र, and ह
iv. ट, द, andप
v. झ, त, फ, and भ
vi. उ, ऐ, ख, छ, ढ, थ, ब, य, व, स, and ष

Phonemes ङ and ठ do not show any similarity w.r.t. to function coefficients with any other phonemes.

For function F4, the values of the function coefficients representing the function ($F_4 = c_{0,4} + c_{1,4}x_1 + c_{2,4}x_2 + ... + c_{n,4}x_{p,}$) comes out to be same for the phonemes

i. उ and क
ii. आ, ऊ, ए, ज, and ष
iii. झ and ठ
iv. प,फ,भ,श, andह
v. अ, ई, ऐ,औ, ख, छ, ट, ढ, त, थ, न, ब, र, ल, व, and क्ष
vi. इ, ओ, घ, ण, च, द, ध, म, य, and स

Phonemes ङ and ग do not show any similarity w.r.t. to function coefficients with any other phonemes.

For function F5, the values of the function coefficients representing the function ($F_5 = c_{0,5} + c_{1,5}x_1 + c_{2,5}x_2 + ... + c_{n,5}x_{p,}$) comes out to be same for the phonemes

i. अ, उ, ओ, झ, and ट
ii. आ, घ, ठ, and ध
iii. इ, ऊ, औ, छ, ज, थ, प, भ, म, व, and ह
iv. ए, च, द, ब, य, र, श,and ष
v. ई, ऐ, ख, ण, ढ, त, न, फ, ल, स,and क्ष

Phonemes क, ङ, and ग do not show any similarity w.r.t. to function coefficients with any other phonemes

For function F6, the values of the function coefficients representing the function ($F_6 = c_{0,6} + c_{1,6}x_1 + c_{2,6}x_2 + ... + c_{n,6}x_{p,}$) comes out to be same for the phonemes

i. घ, ज, द, प, म, and श
ii. आ, इ, ई, उ, ए, ऐ, औ, ख, ट, ढ, थ, ध, न, फ, ब, भ, य, र, ल, स, and क्ष
iii. अ, ऊ, ओ, ण, च, छ, त, व, ष, and ह

Phonemes क, ग, ङ, झ, and ठ do not show any similarity w.r.t. to function coefficients with any other phonemes.

For function F7, the values of the function coefficients representing the function ($F_7 = c_{0,7} + c_{1,7}x_1 + c_{2,7}x_2 + ... + c_{n,7}x_{p,}$) comes out to be same for the phonemes

i. ठ and ङ
ii. आ, ए, त, र, ष, and ह
iii. क, ट, फ, and म
iv. उ and ज
v. ई, ऐ, ओ, ख, घ, थ, द, य, and व
vi. इ, ऊ, औ, ण, च, छ, ढ, ध, न, प, ब, भ, व, स, श, and क्ष

Phonemes झ and ग do not show any similarity w.r.t. to function coefficients with any other phonemes.

For function F8, the values of the function coefficients representing the function ($F_8 = c_{0,8} + c_{1,8}x_1 + c_{2,8}x_2 + ... + c_{n,8}x_{p,}$) comes out to be same for the phonemes

i. उ and क
ii. ठ, द, ध, फ, ल, and स
iii. ओ, औ, ग, ण, ट, ब, and र
iv. झ, म, ह, and न
v. आ, इ, ख, छ, न, and प
vi. ढ and श
vii. अ, ई, ऊ, ए, ऐ, घ, च, ज, त, थ, भ, य, व, and क्ष

Phoneme ङ do not show any similarity w.r.t. to function coefficients with any other phonemes.

For function F9, the values of the function coefficients representing the function $F_9 = c_{0,9} + c_{1,9}x_1 + c_{2,9}x_2 + ... + c_{n,9}x_{p,}$ comes out to be same for the phonemes

i. ट and श
ii. आ, ऊ, ऐ, औ, ज, ण, त, थ, द, ध, न, ब, ष,and क्ष
iii. अ, उ, झ, and फ
iv. इ, ई, ओ, च, प, म, and ह
v. ग and ङ
vi. ए, ख, छ, ठ, ढ, भ, य, र, ल, व, and स

Phoneme घ and क do not show any similarity w.r.t. to function coefficients with any other phonemes.

For function F10, the values of the function coefficients representing the function ($F_{10} = c_{0,10} + c_{1,10}x_1 + c_{2,10}x_2 + ... + c_{n,10}x_{p,}$) comes out to be same for the phonemes

i. आ and ठ
ii. अ, ऊ, ऐ,औ, छ, ढ, ण, त, थ, न, फ, भ, य, ल, व, स, ष, and क्ष
iii. घ, ट, and ड
iv. ए, ज, and प
v. ई, उ, ओ, च, ध, ब, म, and र
vi. इ, ख, झ, द, श, and ह

Phoneme क and ग do not show any similarity w.r.t. to function coefficients with any other phonemes.

Table 5.6 shows the component functions (F11-F20) of the mapping for male to male (M1-M2) transformation. For function F11, the values of the function coefficients representing the function $(F_{11} = c_{0,11} + c_{1,11}x_1 + c_{2,11}x_2 + ... + c_{n,11}x_{p,})$ comes out to be same for the phonemes

i. इ, ई, क, ज, त, and श
ii. झ and ह
iii. अ, आ, च, ठ, and म
iv. उ, ऊ, ए, ऐ, ओ, ढ, प, and र
v. ख, ट, द, ब, भ, and ल
vi. औ, छ, ण, थ, ध, न, फ, य, व, स, and क्ष

Phoneme डand ग do not show any similarity w.r.t. to function coefficients with any other phonemes.

For function F12, the values of the function coefficients representing the function $(F_{12} = c_{0,12} + c_{1,12}x_1 + c_{2,12}x_2 + ... + c_{n,12}x_{p,})$ comes out to be same for the phonemes

i. क and ह
ii. उ, ग, and श
iii. ख, घ, झ, ठ, ड, and ल
iv. आ, ई, ओ, च, छ, ण, त, and न
v. प, म, and र
vi. अ, इ, थ, द, फ, and स
vii. ऊ, ए, ऐ, ज, ट, ढ, ध, ब, य, व ,ष, and क्ष

Phoneme ओ do not show any similarity w.r.t function coefficients with any other phonemes.

For function F13, the values of the function coefficients representing the function $(F_{13} = c_{0,13} + c_{1,13}x_1 + c_{2,13}x_2 + ... + c_{n,13}x_{p,})$ comes out to be same for the phonemes

i. उ, द, and फ
ii. इ, झ, ठ, प,and म
iii. आ, ट, थ, ब, य, स,and श
iv. क and ड
v. ई, ए, ज, and र
vi. ओ, ख, घ, च, and ह
vii. अ, ऊ, ऐ, औ, छ, ढ, ण, त, ध, न, भ, ल, व, and ष

Phoneme ग do not show any similarity w.r.t. to function coefficients with any other phonemes.

For function F14, the values of the function coefficients representing the function $(F_{14} = c_{0,14} + c_{1,14}x_1 + c_{2,14}x_2 + ... + c_{n,14}x_{p,})$ comes out to be same for the phonemes

i. अ, घ, and झ
ii. ख, छ, ण, न, भ, and र
iii. क, ड, and ह
iv. इ, उ, च, ट, त, ध, श, and ष
v. ठ and प
vi. आ, ई, ऊ, ए, ऐ, ओ, औ, ढ, थ, द, ब, म, य, ल ,व, स, and क्ष

Phonemes ज and ग do not show any similarity w.r.t. to function coefficients with any other phonemes

For function F15, the values of the function coefficients representing the function $(F_{15} = c_{0,15} + c_{1,15}x_1 + c_{2,15}x_2 + ... + c_{n,15}x_{p,})$ comes out to be same for the phonemes

i. ट and श
ii. झ and ड
iii. उ, ओ, घ, त, द, ब,and ल
iv. इ, ई, ऐ, ख, ढ, ण, थ, फ, र,and स
v. अ, च, ज, and म
vi. ग and ठ
vii. आ, ऊ, ए, ओ, छ, ध, न, प, भ, य, व, ष, and क्ष

Phoneme क do not show any similarity w.r.t. to function coefficients with any other phonemes

For function F16, the values of the function coefficients representing the function $(F_{16} = c_{0,16} + c_{1,16}x_1 + c_{2,16}x_2 + ... + c_{n,16}x_{p,})$ comes out to be same for the phonemes

i. इ, ट, ठ, and ह
ii. झ and ड
iii. अ, उ, ऊ, ए, क, ख, ढ, भ, र, ल, स, and ष
iv. ज, ण, द, ध, प, ब, म,and व
v. ओ and श
vi. आ, ई, ए, औ, च, छ, त, थ, न, फ, य, and क्ष

Phonemes ग and घ do not show any similarity w.r.t. to func-

tion coefficients with any other phonemes.

For function F17, the values of the function coefficients representing the function ($F_{18} = c_{0,18} + c_{1,18}x_1 + c_{2,18}x_2 + ... + c_{n,18}x_{p,}$) comes out to be same for the phonemes

i. उ, घ, फ, र, ष, and ह
ii. च and ज
iii. अ, आ, ई, ऐ, ओ, औ, छ, त, द, ध, भ, म, व, स, and क्ष
iv. झ and श
v. इ, ऊ, ए, ख, ट, ठ, ड, ढ, ण, थ, न, ब, य, and ल

Phonemes क, ग, and प do not show any similarity w.r.t. to function coefficients with any other phonemes.

For function F18, the values of the function coefficients representing the function ($F_{18} = c_{0,18} + c_{1,18}x_1 + c_{2,18}x_2 + ... + c_{n,18}x_{p,}$) comes out to be same for the phonemes

i. घ and ड
ii. च and ज
iii. क, ठ, and श
iv. ऐ, ओ, औ, ख, झ, प, ब, र, स, and ष
v. अ, आ, इ, ई, उ, ऊ, ए, छ, ढ, ण, त, थ, द, ध, न, फ, भ, म, य, ल, व, and क्ष

Phonemes ग, ह, and ट do not show any similarity w.r.t. function coefficients with any other phonemes

For function F19, the values of the function coefficients representing the function ($F_{19} = c_{0,19} + c_{1,19}x_1 + c_{2,19}x_2 + ... + c_{n,19}x_{p,}$) comes out to be same for the phonemes

i. उ, ख, ट, ठ, and स
ii. घ and झ
iii. अ, इ, ई, ए, ओ, औ, ण, च, छ, ढ, त, थ, ध, न, ब, भ, र, ल, and क्ष
iv. प, फ, श, and ष
v. अ, ग, and ज
vi. आ, ऊ, ऐ, द, म, य, व, and ह

Phonemes क and ड do not show any similarity w.r.t. to function coefficients with any other phonemes.

For function F20, the values of the function coefficients representing the function ($F_{20} = c_{0,20} + c_{1,20}x_1 + c_{2,20}x_2 + ... + c_{n,20}x_{p,}$) comes out to be same for the phonemes

i. उ and घ
ii. ठ, श, and ह
iii. आ, इ, ई, ऊ, ए, ऐ, ओ, औ, ख, ट, थ, द, ध, न, ब, भ, र, व, स, and क्ष
iv. घ, ज, and ष
v. झ and ड
vi. अ, ण, छ, ढ, त, प, फ, म, य, and ल

Phonemes क and ग do not show any similarity w.r.t. function coefficients with any other phonemes.

From the above analysis, it is clear that the20 component functions of the transformation functions for different phonemes may be grouped into eight classes. Further, very few phonemes come in C1, C2, C3, C5, and C7 group and maximum phonemes comes in C6 and C8 group.

As demonstrated in this document, the numbering for sections upper case Arabic numerals, then upper case Arabic numerals, separated by periods. Initial paragraphs after the section title are not indented. Only the initial, introductory paragraph has a drop cap.

4 Conclusion

Investigation was carried out to study the similarity of the transformation functions for different phonemes. All the phonemes are divided into 8 classes' and class dependent function errors were estimated. Investigation was carried out with four speaker pairs. In case of male to male transformation (M1-M2), maximum phonemes come in C6 and C8 group. For male to female transformation (M4-F4) maximum phonemes come in C4, C6, C7, and C8 group.In case of female to male transformation (F3-M3), maximum phonemes come in C4, C6, C7, and C8 group and for female to female transformation (F1-F2) maximum phonemes come in C4, C6, C7, and C8 group. Investigations of the results show that the group can be further reduced to four classes.

Investigations of the results show that the group can be further reduced to four classes

REFERENCES

[1] E. Moulines and Y. Sagisaka, Eds., Speaker Transformation State of the Art and Perspectives. Netherlands: Elsevier, 1995.

[2] H. Mizuno and M. Abe, "Voice conversion based on piecewise linear conversion rules of formant frequency and spectrum tilt," in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 1994, vol. 1, pp. 469–472.

[3] L. M. Arslan, "Speaker transformation algorithm using segmental codebooks (STASC)," Speech Commun., no. 28, 1999.

[4] P. K .Lehana, P. C. Pandey, "Transformation of short-term spectral envelope of speech signal using multivariate polynomial modelling," in Proc. National Conf. Commun., 2011, pp. 1-5.

[5] H. Valbret, E. Moulines, and J. P. Tubach, "Voice transformation using PSOLA technique," Speech Commun., vol. 1, pp. 145–148, 1992.

[6] D. Rentzos, S. Vaseghi, Q. Yan, and C. H. Ho, "Voice conversion through

transformation of spectral and intonation features," in Proc IEEE Int. Conf. Acoust., Speech, Signal Process., 2004, vol. 1, pp. 21–24.

[7] Z. W. Shuang, R. Bakis, S. Shechtman, D. Chazan, and Y. Qin, "Frequency warping based on mapping formant parameters," in Proc. Int. Conf. Spoken Lang. Process., 2006.

[8] N. Iwahashi and Y. Sagisaka, "Speech spectrum transformation by speaker interpolation," in Proc. IEEE Int. Conf. Acoust., Speech, Signal, Process., 1994, vol. 1, pp. 461–464.

[9] M. Narendranath, H. A. Murthy, S. Rajendran, and B. Yegnanarayana, "Transformation of formants for voice conversion using artificial neural networks," Speech Commun., vol. 16, no. 2, pp. 207–216, 1995.

[10] Y. Stylianou, "Harmonic plus noise models for speech, combined with statistical methods, for speech and speaker modification," Ph.D. dissertation, Paris, France, 1996.

[11] Y. Stylianou, O. Cappé, and E. Moulines, "Continuous probabilistic transform for voice conversion," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 1998, vol. 6, pp.131–142.

[12] K. Tokuda, T. Kobayashi, and S. Imai, "Speech synthesis using HMMS with dynamic features," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 1996, pp. 389–392.

[14] M. Abe, S. Nakamura, K. Shikano, and H. Kuwabara, "Voice conversion through vector quantization," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 1998, pp. 565-568.

[15] S. Furui, "Research on Individuality features in speech waves and automatic speaker recognition techniques," Speech Commun., vol. 5, no.2, pp. 183-197, 1986.

[16] K. Itoh and S. Saito, "Effects of acoustical feature parameters of speech on perceptual identification of speaker," Trans. IECE, J65-A, pp. 101-108, 1982.

[17] H. Kuwabara and Y. Sagisaka, "Acoustic characteristics of speaker individuality: control and conversion," Speech Commun., vol. 16, pp. 165-173, 1995.

[18] H. Matsumoto, S. Hiki, T. Sone, and T. Nimura, "Multidimensional representation of personal quality of vowels and its acoustical correlates," Trans. IEEE AU, AU-21, pp. 428-436, 1973.

[19] B. F. Necioglu, M.A. Clements, T. P. Barnwell III, and A. Schmidt-Nielsen, "Perceptual relevance of objectively measured descriptors for speaker characterization," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., vol. 2, pp. 869-872, 1998.

[20] J. Makhoul, "Linear prediction: A tutorial review," in Proc. IEEE, 1975, vol. 63, pp. 561-580.

[21] F. Itakura, "Line spectrum representation of linear predictor coefficients of Speech signals," J. Acoust. Soc. Amer., vol. 57, S35 (A), 1975.

[22] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust., Speech, Signal Process., vol. 28, pp. 357-366, 1980.

[23] W. Holmes, J. Holmes, and M. Judd, "Extension of the bandwidth of the JSRU parallel-formant synthesizer for high quality synthesis of male and female Speech," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 1990, vol. 1, pp. 313-316.

[24] R. J. McAulay and T. F. Quatieri, "Sinusoidal coding", in Kleijn and Paliwal (eds.), Speech Coding And Synthesis, pp. 121-173, B.V. Elsevier Science, Netherlands, 1995.

[25] D. G. Childers and C.-K. Lee, "Vocal quality factors: analysis, synthesis, and perception," J. Acoust. Soc. Amer., pp. 2394-2410, 1990.

[26] D. G. Childers, "Glottal source modeling for voice conversion," Speech Commun., vol. 16, no. 2, pp. 127-138, 1995.

[27] G. Fant, J. Liljencrants and Q. Lin, "A four-parameter model of the glottal flow," Speech transmission laboratory quarterly progress and status reports, No. 4, Royal Institute of Technology, Stockholm, Sweden, pp. 1-13, 1985.

[28] L. R. Rabiner, and R. W. Schafer, "Digital processing of speech signals," Prentice-Hall Inc., Englewood Cliffs. New Jersey, 1978.

[29] D.Talkin, "A robust algorithm for pitch tracking (RAPT)," in Kleijn and Paliwal (eds.), Speech Coding And Synthesis, pp. 121-173, Elsevier Science B.V.Netherlands, 1995.

[30] F. Itakura , "Line spectrum representation of linear predictor coefficients of speech signals," J. Acoust. Soc. Amer., vol. 57, S35 (A), 1975.

[31] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," in Proc. IEEE, vol 77, no.2, pp. 257-286, 1989.

[32] A. Acero, "Acoustical and environmental robustness in automatic speech recognition," Kluwer Academic Publishers, Dordrecht, 1993.

[33] L. M. Arslan, and D. Talkin, "Voice Conversion by Codebook mapping of line spectral frequencies and excitation spectrum," in Proc. Eurospeech, 1997, Rhodes, Greece, vol. 3, pp. 1347-1350.

[34] C. Drioli, "Radial basis function networks for conversion of sound spectra," J. Applied Signal Process., vol. 2001, no. 1, pp. 36-44, 2001.

[35] L. Knohl and A. Rinscheid, "Speaker normalization with self-organizing feature maps," in Proc. Int. Joint Conf. Neural Networks, pp. 243- 246, 1993.

[36] H. Mizuno, and M. Abe, "Voice conversion algorithm based on piecewise linear conversion rules of formant frequency and spectrum tilt," Speech Commun., vol. 16, pp. 153-164, 1995.

[37] J. Laroche, Y. Stylianou, and E. Moulines, "HNS: speech modification based on a harmonic + noise model," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Minneapolis, 1993, vol. 2, pp. 550-553,

[38] W. Endres, W. Bambach, and G. Flosser, "Voice spectrograms as a function of age, voice disguise, and voice imitation," J. Acoust. Soc. Amer., vol. 49, pp. 1842–1848, 1971.

[39] M. R. Sambur, "Selection of acoustic features for speaker identification," IEEE Trans. Acoust. Speech Signal Process., vol. ASSP-23, pp. 176–182, 1975.

[40] P. K. Lehana, "Spectral mapping using multivariate polynomial modelling for voice conversion," Ph. D. thesis, Department of Electrical Engineering, IIT Bombay, 2013.

Table 1(a) component functions(F_1-F_{10}) of the mapping for M1-M2 transformation

Function	Class							
	C1	C2	C3	C4	C5	C6	C7	C8
F_1	क, च	ज	ग	उ, ट, ड, फ, स श, ष	प	ऊ, औ, घ, ण, छ, ठ, ढ, ध, न म, र, ल, व, ह	झ	अ, आ, इ, ई, ओ, ए, ऐ, ख, त, थ, द, ब, भ, य, क्ष
F_2	झ, ठ	उ, द, ल	ड, प, भ, म, व	ई, ओ, च, ज, य	क ग	इ, ए, छ, थ, ब, श, ष, ह	अ, आ, ऊ, ऐ, ओ, ण, न	ख, घ, ट, ढ, त, ध, फ, र, स, क्ष
F_3	क, ग	इ, ई, ओ, औ, ण, च ध ज र ह	ड	अ, आ, उ, ए, घ, ज, र, ह	ठ	ट, द, प	झ, त, फ, भ	उ, ऐ, ख, छ, ढ, थ, ब, य, व, स, ष
F_4	ग	उ, क	आ, ऊ, ए, ज, ष	ड	झ ठ	प, फ, भ, श, ह	अ, ई, ऐ, औ, ख, छ, ट, ढ, त, थ, न, ब, र, ल, व, क्ष	इ, ओ, घ, ण, च, द, ध, म, य, स
F_5	ग	अ, उ, ओ, झ, ट	ड	आ, घ, ठ, ध	क	इ, ऊ, औ, छ, ज, थ, प, भ, म, व, ह	ए, च, द, ब, य, र, श, ष	ई, ऐ, ख,ण, ढ, त, न, फ, ल, स, क्ष
F_6	ग	क	घ, ज, द, प, म, श	आ, इ, ई, उ, ए, ऐ, औ, ख, ट, ढ, थ, ध, न, फ, ब, भ, य, र, ल, स, क्ष	ठ	झ	अ, ऊ, ओ, ण, च, छ, त, व, ष, ह	ड
F_7	ठ, ड	झ	ग	अ, आ, ए, त, र, ष, ह	क, ट, फ, म	उ, ज	ई, ऐ, ओ, ख, घ, थ, द, य, व	इ, ऊ, औ, ण च, छ, ढ, ध, न, प, ब, भ, व, स ,श, क्ष,
F_8	उ, क	ठ, द, ध, फ, ल, स	ड	ओ, औ, ग, ण, ट, ब, र	झ, म, ह, न	आ, इ, ख, छ, न, प	ढ, श	अ, ई, ऊ,ए, ऐ, घ, च, ज, त, थ भ, य, व, क्ष
F_9	घ	ट, श	क	आ, ऊ, ऐ, औ, ज, ण, त, थ, द, ध, न, ब, ष, क्ष	अ, उ, झ, फ	इ, ई, ओ, च, प, म, ह	ग, ड	ए, ख, छ, ठ, ढ भ, य, र, ल, व स
F_{10}	ग	आ, ठ	अ, ऊ, ऐ, औ, छ, ढ, ण, त, थ, न, फ, भ, य, ल, व, स, ष, क्ष	क	घ, ट, ड	ए, ज, प	ई, उ, ओ, च, ध, ब, म, र	इ, ख, झ, द, श, ह

Table 1(b) component functions(F_{11}-F_{20}) of the mapping for M1-M2 transformation

Function	Class							
	C1	C2	C3	C4	C5	C6	C7	C8
F_{11}	ग	इ, ई, क, ज, त श	घ, झ, ह	अ, आ, च, ठ, म	ड	उ, ऊ, ए, ऐ, ओ, ढ, प, र	ख, ट, द, ब, भ, ल, औ, छ	औ, छ, ण, थ, ध, न, फ, य, व, स, क्ष
F_{12}	क, ह	उ, ग, श	ख, घ, झ, ठ, ड, ल	आ, ई, ओ, च, छ, ण, त, न	ओ	प, म, र	अ, इ, थ, द, ,फ स	ऊ, ए, ऐ, ज, ट, ढ, ध, ब, य, व, ष, क्ष
F_{13}	ग	उ, द, फ	इ, झ, ठ, प, म	आ, ट, थ, ब, य, स, श	क, ड	ई, ए, ज, र	ओ, ख, घ, च, ह	अ, ऊ, ऐ, औ, छ, ढ, ण, त, ध, न, भ, ल, व, ष
F_{14}	ज	ग	अ, घ, झ	ख, छ, ण, न, भ, र	क, ड, ह	इ, उ, च, ट, त, ध, श, ष	ठ, प	आ, ई, ऊ, ए, ऐ, ओ, औ, ढ, थ, द, ब, म, य, ल, व, स, क्ष
F_{15}	ट, श	झ, ड	क	उ, ओ, घ, त, द, ब, ल	इ, ई, ऐ, ख, ढ ण, थ, फ, र, स	अ, च, ज, म	ग, ठ	आ, ऊ, ए, ओ, छ, ध, न, प, भ, य, व, ष, क्ष
F_{16}	ग	इ, ट, ठ, ह	झ, ड	अ, उ, ऊ, ए, क, ख, ढ, भ, र, ल, स, ष	घ	ज, ण, द, ध, प, ब, म, व	ओ, श	आ, ई, ए, औ, च, छ, त, थ, न, फ, य, क्ष
F_{17}	ग	उ, घ, फ, र, ष, ह	प	क	च, ज	अ, आ, ई, ऐ, ओ, औ, छ, त, द ,ध, भ, म, व स, क्ष	झ, श	इ, ऊ, ए, ख, ट, ठ, ड, ढ, ण, थ, न ,ब, य, ल
F_{18}	ग	घ, ड	च, ज	क, ठ, श	ह	ऐ, ओ, औ, ख, झ, प, ब, र, स, ष	ट	अ, आ, इ, ई, उ, ऊ, ए, छ, ढ, ण, त, थ, द, ध, न, फ, भ, म, य, ल, व,क्ष
F_{19}	उ, ख, ट, ठ, स	ड	क	घ, झ	अ, इ, ई, ए, ओ, औ, ण, च, छ, ढ, त, थ, ध, न, ब, भ, र, ल, क्ष	प, फ, श, ष	ग, ज	आ, ऊ, ऐ, द, म, य, व, ह
F_{20}	उ, घ	क	ग	ठ, श, ह	आ, इ, ई, ऊ, ए, ऐ, ओ, औ, ख, ट, थ, द, ध, न, ब, भ, र, व, स, क्ष	घ, ज, ष	झ, ड	अ, ण, छ, ढ, त, प, फ, म, य ल