Automatic speech pronunciation correction with dynamic frequency warping-based spectral conversion

Nobukatsu Hojo, Hirokazu Kameoka, Kou Tanaka, Takuhiro Kaneko
NTT Communication Science Laboratories, NTT Corporation, Japan
hojo.nobukatsu@lab.ntt.co.jp

***Abstract*—This paper deals with the problem of pronunciation conversion (PC) task, a problem to reduce non-native accents in speech while preserving the original speaker identity. Although PC can be regarded as a special class of voice conversion (VC), a straightforward application of conventional VC methods to a PC task would not be successful since with VC the original speaker identity of input speech may also change. This problem is due to the fact that two functions, namely an accent conversion function and a speaker similarity conversion function, are entangled in an acoustic feature mapping function. This paper proposes dynamic frequency warping (DFW)-based spectral conversion to solve this problem. The proposed DFW-based PC converts the pronunciation of input speech by relocating the formants to the corresponding positions in which native speakers tend to locate their formants. We expect the speaker identity is preserved because other factors such as formant powers are kept unchanged. in a low frequency domain evaluation results confirmed that DFW-based PC with spectral residual modeling showed higher speaker similarity to original speaker while showing a comparable effect of reducing foreign accents to a conventional GMM-based VC method.**

***Index Terms*—Accent conversion, dynamic frequency warping, voice conversion**

I. Introduction

This paper deals with the problem of pronunciation conversion (PC), a problem to reduce accent in speech while preserving the speaker identity of the original speech. Accents are differences in pronunciation by a community of people from a national or regional geographical area, or a social grouping [1]. It is known that differences by accents are manifested in the differences such as the formants and their trajectories [2], [3] or pitch intonation and duration parameters [4], [1]. Reducing these accents from speech while preserving the original speaker identity will be beneficial for applications such as language educational systems for second language learners [5], [6], [7] and teleconference scenarios where people with different nationalities participate.

If we consider an accent as one of non/para-linguistic information, PC can be regarded as a special class of a voice conversion (VC) problem, a problem to convert non/para-linguistic information while preserving linguistic information. For VC, data-driven statistical methods have been successfully introduced during the last two decades [8], [9], [10]. One successful VC method is based on Gaussian mixture models (GMMs) [9], [10]. Most of the conventional VC methods first train a mapping function between the acoustic features of source and target speech using parallel data, i.e. a pair of time-aligned feature sequences of source and target speech. At test time, a feature sequence of the input speech is converted using the trained mapping function. A direct application of these methods to the PC task would not be successful because the original speaker identity may also change. This problem is because two functions, the accent conversion function and the speaker identity conversion function, are entangled in an acoustic feature mapping function.

This work is based on a belief that a dynamic frequency warping (DFW)-based spectral conversion approach can be a reasonable solution to this problem. The DFW-based spectral conversion approach was originally proposed [11], [12] mainly for the purpose of improving naturalness of converted speech. Since one of the dominant differences in accented speech appears in formant frequency trajectories [2], [3], we expect that pronunciation can be corrected via frequency warping if the formants can be relocated to the positions in which native speakers tend to locate their formants. In this way, we expect that the original speaker identity will not be affected since DFW does not convert the other factors such as formant powers and spectral tilts.

The main purpose of this work is to investigate the effectiveness of DFW-based spectral conversion for the PC task. Furthermore, there are two points that we want to investigate. One is the effectiveness of spectral power interpolation. DFW only allows us to deform source spectra in the frequency direction and does not have an ability to reduce the gap in spectral powers. Since the peakiness of each formant is also expected to be an important factor that characterizes pronunciation quality, we would also want to deform source spectra in the power direction. However, if we convert source spectra exactly to target spectra, the speaker identity will no longer be preserved. Thus, there is a trade-off between the pronunciation quality and the voice similarity to a source speaker. This trade-off is investigated by subjective evaluation. The other point is concerned with the modeling of frequency warping functions. The idea of the method in [12] is to associate a frequency warping function with each Gaussian of the GMM that models a joint distribution of source and target spectra. The frequency warping function is obtained from a pair of source and target spectra averaged over all the frames assigned to the same Gaussian. We found that the averaged spectra tend to be over-smoothed and so the obtained warping function will also be over-smoothed. To avoid this, we propose a method that first extracts frequency warping functions from pairs of source and target spectra frame-by-frame, treats the obtained frequency warping functions as features to be predicted, and models the joint distribution of source spectra and the frequency warping

ISBN 978-90-827970-1-5 © EURASIP 2018

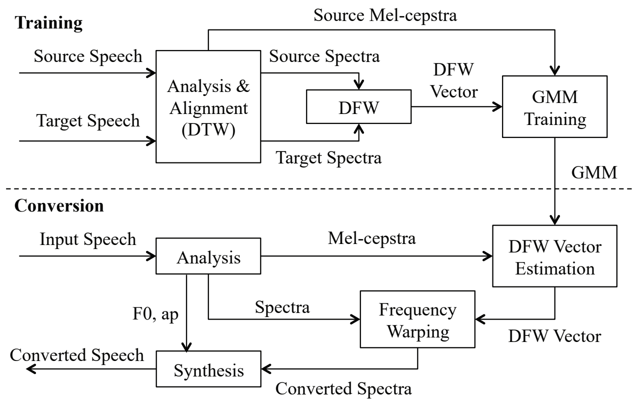


Fig. 1. Architecture of the proposed accent conversion method.

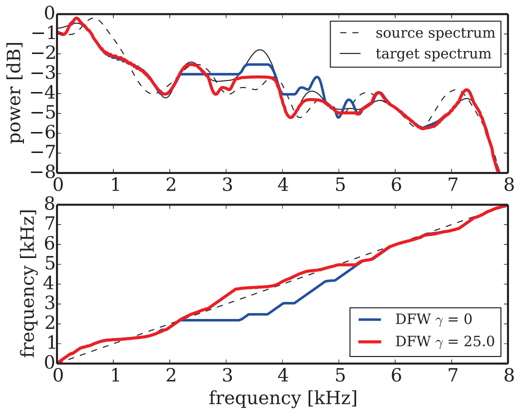


Fig. 2. Example of DFW spectra (top) and extracted DFW vector (bottom). The blue and red line shows the result using $\gamma = 0$ and $\gamma = 25.0$ in eq. (2), respectively.

functions using a GMM. In this paper, we particularly focus on the problem of spectral conversion only. However, it can be used in combination with prosody conversion methods [5], [13].

II. Dynamic Frequency Warping-based Accent Conversion

The proposed method consists of training process and conversion process. We show the overall architecuture of the proposed method in Fig. 1.

A. Dynamic frequency warping with frequency derivative distance

The proposed method first finds an optimal warping of frequency axis with DFW [14], [15]. Let us denote the time aligned source spectra by $\boldsymbol{X} = \{\boldsymbol{x}_t\}_{t=1}^T$ and the target spectra by $\boldsymbol{Y} = \{\boldsymbol{y}_t\}_{t=1}^T$. Here $\boldsymbol{x}_t = [x_{f,t}]_{f=1}^F$ and $\boldsymbol{y}_t = [y_{f,t}]_{f=1}^F$ respectively denote the source and target spectrum. t, f denote the frame and frequency indices, respectively. The warping function $\hat{\boldsymbol{w}}_t = [\hat{w}_{f,t}]_{f=1}^F$, which we call the DFW vector, can be obtained as the path that minimizes the spectral distance between a frequency warped source spectrum and a target spectrum;

$$\hat{\boldsymbol{w}}_t = \underset{w_1,\cdots,w_F}{\arg\min} \sum_{f=1}^{F} \mathcal{D}(x_{w_f,t}, y_{f,t}), \tag{1}$$

where w_f takes a frequency index $w_f \in \{1, \cdots, F\}$ in an ascending order. By restricting the warping path to be $w_{f+1} - w_f \in \{0, 1, 2\}$ for each $f \in \{1, \cdots, F-1\}$, a smoothed and fixed length DFW vector can be extracted. Although it is possible to use the l_2 norm of a log spectral difference to define $\mathcal{D}(x, y)$, the obtained warped spectra tend to have plateau because of the power difference of spectral peaks between source and target speakers. We concern that the plateau can degrade harmonics and quality of converted speech, which is the reason we define the distance introducing the frequency derivative distance term as follows in this work;

$$\begin{aligned} \mathcal{D}(x_f, y_f) =& || \log x_f - \log y_f ||_2 \\ &+ \gamma || \dot{x}_f - \dot{y}_f ||_2, \end{aligned} \tag{2}$$

$$\dot{x}_f = \log x_{f+1} - \log x_f, \tag{3}$$

$$\dot{y}_f = \log y_{f+1} - \log y_f \tag{4}$$

where and γ indicates the weight for the frequency derivative term. This simple spectral distance is expected to have a similar effect to correlation based DFW [16] or DFW based on the histogram of spectral peaks [17]. An example of spectral warping using frequency derivative distance is illustrated in Fig.2. We can see that the spectral plateau is eliminated by introducing the frequency derivative distance term.

Since we would want to eliminate the speaker identity information from the DFW vector as much as possible, we found it necessary to apply vocal tract length normalization (VTLN) [18] to the source and target speech spectra before performing DFW.

B. GMM training and DFW vector estimation

For estimating DFW vectors from the source spectra, we model the joint distribution of the source spectra and the DFW vectors using a GMM. In order to avoid overfitting in modeling high dimensional vectors, we compress the dimension of the feature vectors and constitute a joint vector $\boldsymbol{z}_t = [\boldsymbol{m}_t^{\mathrm{T}}, \boldsymbol{d}_t^{\mathrm{T}}]^{\mathrm{T}}$, where $\boldsymbol{m}_t$ is the mel-cepstrum extracted from the source spectrum $\boldsymbol{x}_t$ at frame t. The vector $\boldsymbol{d}_t$ is derived as follows;

$$\boldsymbol{d}_t = \mathrm{DCT}(\boldsymbol{w}_t - \boldsymbol{w}_b), \tag{5}$$

where $\boldsymbol{w}_b = [1, \cdots, F]^{\mathrm{T}}$ and $\mathrm{DCT}(\cdot)$ denotes the discrete cosine transform. We model the joint feature as follows;

$$p(\boldsymbol{z}) = \sum_{i=1}^{I} \alpha_i N(\boldsymbol{z}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \sum_{i=1}^{I} \alpha_i = 1, \alpha_i > 0, \tag{6}$$

where $N(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the normal distribution with the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$. α_i denotes a weight of class i, and I denotes the total number of the Gaussian mixtures. Given a training set of $\{\boldsymbol{m}_t, \boldsymbol{d}_t\}_{t=1}^T$ pairs, we train the GMM parameters using the EM algorithm.

ISBN 978-90-827970-1-5 © EURASIP 2018

At test time, DFW vectors are estimated using the trained GMM and the input spectra. The mapping function [10] is given by

$$\begin{aligned} F(\boldsymbol{m}) &= E[\boldsymbol{d}|\boldsymbol{m}] \\ &= \sum_{i=1}^{I} h_i(\boldsymbol{m})[\boldsymbol{\mu}_i^{(d)} + \boldsymbol{\Sigma}_i^{(dm)}\big(\boldsymbol{\Sigma}_i^{(mm)}\big)^{-1}(\boldsymbol{m} - \boldsymbol{\mu}_i^{(m)})] \end{aligned} \tag{7}$$

$$h_i(\boldsymbol{m}) = \frac{\alpha_i N\big(\boldsymbol{m}; \boldsymbol{\mu}_i^{(m)}, \boldsymbol{\Sigma}_i^{(mm)}\big)}{\sum_j^I \alpha_j N\big(\boldsymbol{m}; \boldsymbol{\mu}_j^{(m)}, \boldsymbol{\Sigma}_j^{(mm)}\big)}, \tag{8}$$

where $\boldsymbol{\mu}_i^{(m)}$ and $\boldsymbol{\mu}_i^{(d)}$ denote the mean vectors of class i for the mel-cepstra and DCT of DFW vectors. The estimated DFW vector is derived as follows;

$$\tilde{\boldsymbol{w}}_t = \mathrm{iDCT}\big(F(\boldsymbol{m}_t)\big) + \boldsymbol{w}_b, \tag{9}$$

where $\mathrm{iDCT}(\cdot)$ denotes the inverse discrete cosine transform. Since the estimated $\tilde{\boldsymbol{w}}_t$ takes a continuous value, they are floored to be an integer frequency index to convert spectra as $\tilde{\boldsymbol{y}}_t = [x_{\tilde{w}_{f,t},t}]_{f=1}^F$.

C. Spectral residual modeling

Note that DFW only has an ability to deform source spectra in the frequency direction and does not have an ability to fill the gap in spectral powers. Since the power of each formant is also expected to be an important factor that characterizes pronunciation quality, we would also want to deform source spectra in the power direction. However, if we convert source spectra exactly to target spectra, the speaker identity will no longer be preserved. Thus, there is a trade-off between the pronunciation quality and the voice similarity to a source speaker. Here, we also consider predicting the power differences between the warped spectra and the target spectra so that we can add the predicted differences to the warped spectra at test time.

The residual spectra $\boldsymbol{r}_t = [r_{f,t}]_{f=1}^F$ is defined as the difference between the target spectra and the warped source spectra as follows;

$$r_{f,t} = \frac{y_{f,t}}{x_{\hat{w}_{f,t},t}} \tag{10}$$

This residual spectra is extracted for each frame and construct a joint vector $\boldsymbol{s}_t = [\boldsymbol{m}_t^T, \boldsymbol{q}_t^T]^{\mathrm{T}}$, where $\boldsymbol{q}_t$ denotes the DCT of $\log \boldsymbol{r}_t$. This joint vector is modeled by another GMM and used to estimate residual spectra $\tilde{\boldsymbol{r}}$ in the same manner for estimating the residual vector. The output converted spectra $\tilde{\boldsymbol{y}}_t^{(r)} = [\tilde{y}_{f,t}^{(r)}]_{f=1}^F$ is derived as follows;

$$\tilde{y}_{f,t}^{(r)} = \tilde{y}_{f,t} \cdot \tilde{r}_{f,t}^{\lambda} \tag{11}$$

where λ denotes the weight for spectral residual modeling.

III. Experimental Evaluation

A. Experimental conditions

We evaluated pronunciation similarity and speaker identity of the converted speech to compare the performance of the conventional VC and proposed methods. We used an Indian male speaker as the source (hereafter, the SRC) and an American male speaker as the target (the TGT). The dataset consisted of 65 pair utterances (5.2 minutes). We used 20 utterances for evaluation and the others for training. Table. I shows the statistical information of the corpus. The data were sampled at 16kHz, then 25 mel-cepstral coefficients, fundamental frequency (F_0), and aperiodicities were extracted every 5ms by using the STRAIGHT analysis system [19]. To obtain parallel utterances, we used dynamic time warping (DTW) to align mel-cepstral sequences of the source and target speakers [20]. We evaluated and compared the following 6 spectral conversion methods.

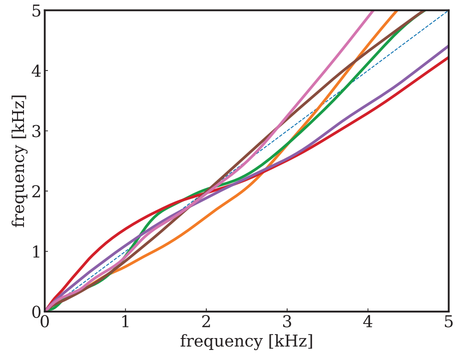


Fig. 3. Example of DFW vectors reconstructed from $\boldsymbol{\mu}_i^{(d)}$. Six out of 16 Gaussian components are shown for simplicity.

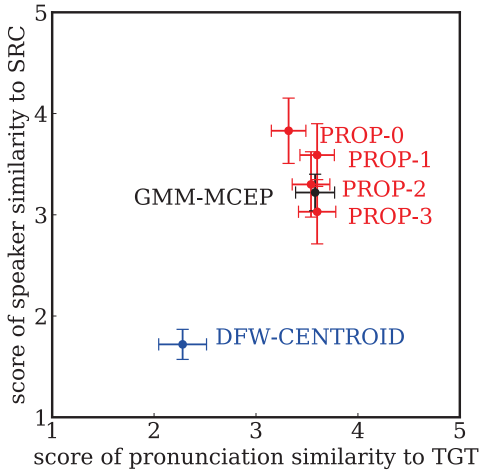
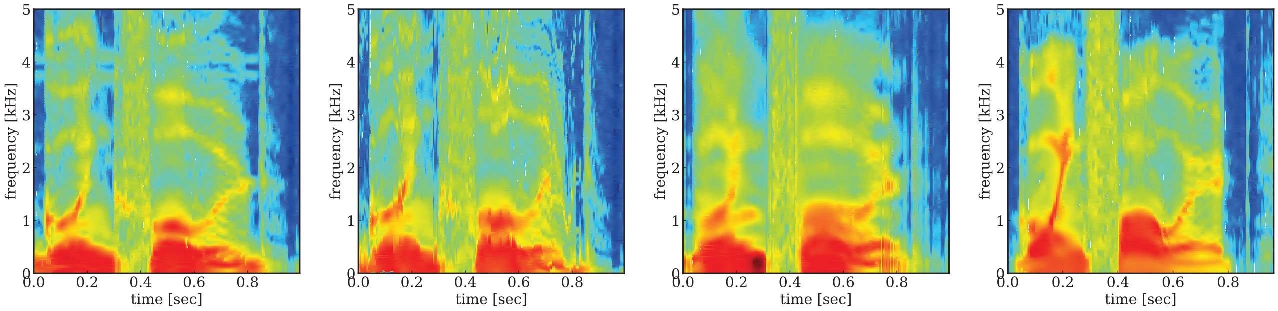


Fig. 4. Subjective evaluation results.

- GMM-MCEP: The conventional GMM-based VC method using mel-cepstral feature s [10].
- DFW-CENTROID: DFW-based spectral conversion method. DFW functions were derived from centroid spectra pair of each Gaussian.
- PROP-0: The proposed method without spectral residual modeling ($\lambda = 0.00$).
- PROP-1: The proposed method with spectral residual modeling ($\lambda = 0.33$).
- PROP-2: The proposed method with spectral residual modeling ($\lambda = 0.67$).

ISBN 978-90-827970-1-5 © EURASIP 2018



(a) SRC (b) PROP-1 (c) GMM-MCEP (d) TGT

Fig. 5. Example of spectrogram of SRC, PROP-1, GMM-MCEP and TGT for a utterance "Going forward.". We show frequency region of 0 - 5 kHz to clarify the formant transition.

TABLE I
STATISTICAL INFORMATION OF THE SPEECH CORPUS

Speakers	2 males
Total speech length	5.2 minutes
Number of utterances	65 sentences
Average number of words per sentence	5.8 words

- PROP-3: The proposed method with spectral residual modeling ($\lambda = 1.00$).

Although DFW-CENTROID was implemented emulating the conventional DFW-based VC method [12], details such as vocoder for synthesizing speech was not the same. This is in order to reduce the factors considered in the experiments. We evaluated 4 proposed methods with different spectral residual weight in order to evaluate the trade-off of the pronunciation quality improvement and speaker identity degradation.

For DFW-CENTROID, we first train a GMM that models joint distribution of source and target spectra using line spectral pairs (LSP) features. Then, a pair of source and target centroid spectra was reconstructed from the mean vector of each Gaussian. The frequency warping function was obtained from the pair of source and target centroid spectra using the DFW method in sec. II-A. The conversion procedure was similar to the proposed method except for using LSPs instead of mel-cepstra as input spectral features. We used 25 dimensional LSP for this method.

For the proposed methods, we used 25 dimensional mel-cepstra and 25 dimensional DCT of DFW vectors. We used mel-spectra for DFW vector extraction so that the warping function gets higher frequency resolution in a low frequency domain. We set $\gamma = 25.0$ in eq. (2), which was determined experimentally. We show the results without conducting VTLN before DFW vector extraction, because the results were not improved. We consider this is because the vocal tract length of SRC and TGT was almost same.

For all of the methods, the number of mixture components for GMMs was 16. All of the speech samples were synthesized by STRAIGHT vocoder [19], using converted spectra by each method, F0 and aperiodicity of the SRC speaker.

Fig. 3 shows the example of warping functions reconstructed from $\boldsymbol{\mu}_i^{(d)}$ of the trained GMM of PROP-0. We can see that these functions can warp frequency by 200 Hz around $1-3$ kHz frequency region, which we expect is enough warping width to relocate formants to reduce accents.

B. Subjective evaluation

Two subjective evaluations were conducted. We evaluated the pronunciation similarity to the TGT and speaker similarity to the SRC by DMOS tests. The subjects rated the pronunciation and the speaker similarity using a 5–point scale: "5" for absolutely similar, "4" for similar, "3" for fair, "2" for dissimilar "1" for absolutely dissimilar . There were ten participants who were well-educated English speakers.

Fig. 4 shows the pronunciation and speaker similarity scores obtained in the subjective evaluations with confidence intervals of 5%. First we compare the results of GMM-MCEP and PROP-0. We can see that GMM-MCEP shows higher pronunciation similarity score. This result shows that DFW-based spectral warping without spectral residual modeling cannot reduce the foreign accents as much as conventional VC methods. However, if we compare GMM-MCEP and PROP-1, we can see that pronunciation similarity scores were comparable while the speaker similarity score was higher in PROP-1. These results confirmed that DFW-based PC with spectral residual modeling showed higher speaker similarity to original speaker while showing a comparable effect of reducing foreign accents to a conventional GMM-based VC method. We then compare DFW-CENTROID and PROP-0. We can see both the score is improved by PROP-0. These results confirm that the DFW modeling by the proposed method was effective. We then compare the results between the proposed methods. Although the pronunciation similarity score by PROP-1 was higher than PROP-0, those by PROP-2 and PROP-3 were comparable with PROP-1 while the speaker similarity score degrades. These results show that the weight higher than 0.67 was not effective for PC in our experimental conditions.

Fig. 5 shows examples of spectrogram of SRC, PROP-1, GMM-MCEP and TGT for a utterance "Going forward." included in the test set. From these examples, we can see that the formant frequency trajectory around 1 kHz by TGT is located in higher frequency region than SRC. We can also see that this formant frequency trajectory is converted to higher frequency by PROP-1. It is also confirmed that the spectra by PROP-1 is more similar to SRC than GMM-MCEP, which resulted in higher speaker similarity to SRC in our subjective evaluations.

IV. CONCLUSION

This paper investigated the effectiveness of DFW-based spectral conversion for pronunciation conversion task, a prob-

ISBN 978-90-827970-1-5

lem to reduce accent in speech while preserving the speaker identity of the original speech. The proposed method converts the pronunciation of input speech by relocating the formants to the corresponding positions in which native speakers tend to locate their formants. We expect the speaker identity is preserved because other factors such as formant powers are kept unchanged. Subjective evaluation results confirmed that DFW-based pronunciation conversion with spectral residual modeling showed higher speaker similarity to original speaker while showing a comparable effect of reducing foreign accents to a conventional GMM-based VC method. It is worth investigating whether utilizing neural network-based methods [21], [22], [23] for DFW vector estimation is effective with larger size of speech corpus. Future works include extending the experiments to other speaker, accent and language pairs.

References

[1] Qin Yan and Saeed Vaseghi, "A comparative analysis of uk and us english accents in recognition and synthesis," in *Proc. ICASSP 2002*. IEEE, 2002, vol. 1, pp. I–413.

[2] Jonathan Harrington, Felicity Cox, and Zoe Evans, "An acoustic phonetic study of broad, general, and cultivated australian english vowels," *Australian Journal of Linguistics*, vol. 17, no. 2, pp. 155–184, 1997.

[3] Catherine I Watson, Jonathan Harrington, and Zoe Evans, "An acoustic comparison between new zealand and australian english vowels," *Australian journal of linguistics*, vol. 18, no. 2, pp. 185–207, 1998.

[4] John C Wells, *Accents of English*, vol. 1, Cambridge University Press, 1982.

[5] Daniel Felps, Heather Bortfeld, and Ricardo Gutierrez-Osuna, "Foreign accent conversion in computer assisted pronunciation training," *Speech Communication*, vol. 51, no. 10, pp. 920 – 932, 2009, Spoken Language Technology for Education.

[6] Daniel Felps, Christian Geng, and Ricardo Gutierrez-Osuna, "Foreign accent conversion through concatenative synthesis in the articulatory domain," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 8, pp. 2301–2312, 2012.

[7] Daniel Felps and Ricardo Gutierrez-Osuna, "Developing objective measures of foreign-accent conversion," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 5, pp. 1030–1040, 2010.

[8] Masanobu Abe, Satoshi Nakamura, Kiyohiro Shikano, and Hisao Kuwabara, "Voice conversion through vector quantization," *Journal of the Acoustical Society of Japan (E)*, vol. 11, no. 2, pp. 71–76, 1990.

[9] Yannis Stylianou, "Applying the harmonic plus noise model in concatenative speech synthesis," *IEEE Transactions on speech and audio processing*, vol. 9, no. 1, pp. 21–29, 2001.

[10] Tomoki Toda, Alan W Black, and Keiichi Tokuda, "Voice conversion based on maximum-likelihood estimation of spectral parameter trajectory," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 8, pp. 2222–2235, 2007.

[11] Tomoki Toda, Hiroshi Saruwatari, and Kiyohiro Shikano, "Voice conversion algorithm based on gaussian mixture model with dynamic frequency warping of straight spectrum," in *Proc. ICASSP 2001*. IEEE, 2001, vol. 2, pp. 841–844.

[12] Daniel Erro, Asunción Moreno, and Antonio Bonafonte, "Voice conversion based on weighted frequency warping," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 5, pp. 922–931, 2010.

[13] Kwansun Cho and John G Harris, "Towards an automatic foreign accent reduction tool," in *Proc. 3rd Int. Conf. Speech Prosody*, 2006.

[14] Hélene Valbret, Eric Moulines, and Jean-Pierre Tubach, "Voice transformation using psola technique," *Speech communication*, vol. 11, no. 2-3, pp. 175–187, 1992.

[15] Noriyasu Maeda, Banno Hideki, Shoji Kajita, Kazuya Takeda, and Fumitada Itakura, "Speaker conversion through non-linear frequency warping of straight spectrum," in *Sixth European Conference on Speech Communication and Technology*, 1999.

[16] Xiaohai Tian, Zhizheng Wu, Siu Wa Lee, and Eng Siong Chng, "Correlation-based frequency warping for voice conversion," in *Chinese Spoken Language Processing (ISCSLP), 2014 9th International Symposium on*. IEEE, 2014, pp. 211–215.

[17] Elizabeth Godoy, Olivier Rosec, and Thierry Chonavel, "Voice conversion using dynamic frequency warping with amplitude scaling, for parallel or nonparallel corpora," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1313–1323, 2012.

[18] Jordan Cohen, Terri Kamm, and Andreas G Andreou, "Vocal tract normalization in speech recognition: Compensating for systematic speaker variability," *The Journal of the Acoustical Society of America*, vol. 97, no. 5, pp. 3246–3247, 1995.

[19] Hideki Kawahara, Jo Estill, and Osamu Fujimura, "Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system straight," in *Second International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, 2001.

[20] Yannis Stylianou, "Statistical methods for voice quality transformation," *Proc. European Speech Communication Association, Madrid, Spain, 1995*, pp. 447–450, 1995.

[21] Toru Nakashika, Tetsuya Takiguchi, and Yasuo Ariki, "Voice conversion using speaker-dependent conditional restricted boltzmann machine," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2015, no. 1, pp. 8, 2015.

[22] Lifa Sun, Shiyin Kang, Kun Li, and Helen Meng, "Voice conversion using deep bidirectional long short-term memory based recurrent neural networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4869–4873.

[23] Shariq Mobin and Joan Bruna, "Voice conversion using convolutional neural networks," *arXiv preprint arXiv:1610.08927*, 2016.

ISBN 978-90-827970-1-5 © EURASIP 2018