석 사 학 위 논 문
Master's Thesis

가창 표현 이식 알고리즘

Transferring Singing Expressions from One Voice to Another

2017

용 상 언 (龍 相 彦 Yong, Sangeon)

한 국 과 학 기 술 원

Korea Advanced Institute of Science and Technology

석 사 학 위 논 문

가창 표현 이식 알고리즘

2017

용 상 언

한 국 과 학 기 술 원

문화기술대학원

가창 표현 이식 알고리즘

용 상 언

위 논문은 한국과학기술원 석사학위논문으로
학위논문 심사위원회의 심사를 통과하였음

2017년 6월 22일

심사위원장 남 주 한 (인)

심 사 위 원 노 준 용 (인)

심 사 위 원 이 성 희 (인)

Transferring Singing Expressions from One Voice to Another

Sangeon Yong

Advisor: Juhan Nam

A dissertation submitted to the faculty of
Korea Advanced Institute of Science and Technology in
partial fulfillment of the requirements for the degree of
Master of Science in Engineering in Culture Technology

Daejeon, Korea
August 1, 2017

Approved by

Juhan Nam
Professor of Graduate School of Culture Technology

The study was conducted in accordance with Code of Research Ethics[1].

[1] Declaration of Ethical Conduct in Research: I, as a graduate student of Korea Advanced Institute of Science and Technology, hereby declare that I have not committed any act that may damage the credibility of my research. This includes, but is not limited to, falsification, thesis written by someone else, distortion of research findings, and plagiarism. I confirm that my thesis contains honest conclusions based on my own careful research under the guidance of my advisor.

MGCT
20154466

용상언. 가창 표현 이식 알고리즘. 문화기술대학원 . 2017년. 17+iv 쪽. 지도교수: 남주한. (영문 논문)
Sangeon Yong. Transferring Singing Expressions from One Voice to Another. Graduate School of Culture Technology . 2017. 17+iv pages. Advisor: Juhan Nam. (Text in English)

초 록

본 논문에서는 자동적으로 가창 표현을 한 목소리 신호에서 다른 목소리 신호로 이식하는 오디오 신호처리 시스템을 제안한다. 가창자의 능력에 따라 같은 노래를 부르더라도 음의 시작점, 음정, 에너지와 같은 부분에서 큰 변화가 발생할 수 있다. 이 시스템은 이러한 가창자의 고유의 음색을 제외한 음악적인 표현들을 추출 및 적용하는 것에 중점을 두었다. 이러한 가창 표현 이식 행위는 노래 부르기를 어려워하는 사람들의 음악 활동에 도움을 주고, 새로운 가창 표현을 학습하려는 사람들에게 보다 직관적인 가이드라인을 제공해 줄 수 있다. 이 시스템은 차례대로 음의 타이밍 정보와 음정, 그리고 에너지를 일치시키는 방식으로 표현을 이식한다. 본 연구에서는 이를 위해 음의 타이밍 정보를 일치시키는 알고리즘, 음정과 에너지 정보를 일치시키는 알고리즘, 그리고 해당 알고리즘의 성능을 최대한 개선시키고 최적화하는 방법을 제안한다. 그리고 이러한 세부 방법들을 기반으로 가창 표현 이식 시스템을 제안하여 가창 표현 수정에 대한 새로운 접근법을 제시하려고 한다.

핵 심 낱 말 가창, 표현 이식, 시간축 변환, 동적 시간 워핑

Abstract

This paper presents an audio signal processing system that automatically transfers singing expressions from one voice to another. Depending on singers' skills, a song is sung with great variations in terms of note onset time, pitch and energy. The system focused on extracting and transferring musical expressions, excluding the timbre of singers. This singing expression transfer system can provide more intuitive guidance to those who want to learn new vocabulary expressions and help the music activities of those who have difficulty in singing. The system transfers expressions in the order of tempo, pitch, and energy. In this study, we propose an algorithm to align the tempo of the note, a method to match pitch and energy information, and a method to optimize the performance of these processes. Based on these methods, we propose a new singing expression transfer system and propose a new approach to singing voice modification.

Keywords Singing voice, expression transfer, time-scale modification, dynamic time warping

Contents

List of Tables

List of Figures

Chapter 1. Introduction

Singing is a popular musical activity that many people enjoy, for example, in the form of karaoke. Depending on singing skills, a song can be rendered into touching music or just noisy sounds. What if my bad singing can be transformed and so sound like a professional? In this research, we present a vocal processing system that automatically transfers singing expressions from one voice to another.

Commercial vocal correction tools such as Autotune[1], VariAudio[2] and Melodyne[3] mainly focus on modifying pitch of singing voice. Some of them are capable of manipulating note onset timing or other musical expressions by editing transcribed MIDI notes. Although they provide automated controls, the correction process is often tedious and repetitive until satisfactory results are achieved. There are some previous work that attempted to minimize the manual effort in modifying musical expressions. Bryan et. al proposed a variable-rate time-stretching system that allows users to modify the stretching ratio easily [11]. Given a user-guided stiffness curve, the system automatically computed time-dependent stretch rate via a constrained optimization program. Roebel et. al proposed an algorithm to remove vibrato expressions [8]. They operated entirely based on spectral envelope smoothing without manipulation of individual partial parameters. While these methods provide more convenience to process singing voice signals, they still require user guide or parametric control to some extent.

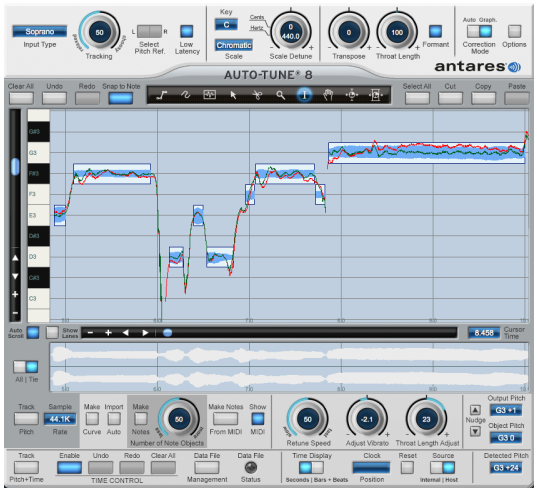

figure 1.1: Antares Autotune 8 Graphical Mode.[5]In the existing vocal correction tools, users should manipulate features in manual to modify singing voice signals.

In this thesis, we propose an audio signal processing system that modifies musical expressions of singing voice in a fully automatic manner with a target singing voice as a control guide. Assuming that both source and target voices sing the same song, the system transfers three musical expressions from target to source: tempo, pitch, and dynamics. First, it temporally synchronizes two singing voices using dynamic time warping on vibrato-suppressed spectrum and a formant feature. Second, it extracts pitch ratio between the two voices and modifies the pitch of source voice using pitch-synchronous overlap-add

[1] http://www.antarestech.com/products/index.php
[2] https://www.steinberg.net/en/products/cubase/cubase_pro.html
[3] http://www.celemony.com/en/melodyne/what-is-melodyne
[5] http://www.antarestech.com/products/detail.php?product=Auto-Tune_8_66

(PSOLA). Finally, it modifies dynamics of the source voice by extracting the ratio of amplitude envelops. In the series of process, the system does not use any user guide or additional information such as lyrics and music scores beside a target voice. Since the system modifies only technical elements in singing and preserves the timbre of source voice, it will be useful for not only sound production but also vocal training.

Chapter 2. Research Background

2.1 Time-Scale Modification Algorithm

Time-scale modification (TSM) algorithm is the process that manipulates the length of the audio signal. [2] The ideal TSM algorithm should modify only the tempo of the signal, and preserve any other properties such as pitch and timbre. This TSM method is commonly used in sound producing area to synchronize the duration of audio sources to other media source, or change the pitch of audio sources without changing the duration of audio sources.

There are two main issues in TSM procedures. The first one is degradation of percussive transients. [19] While modifying audio sources with TSM algorithms, percussive transients often disappeared or are doubled. The other problem is phase discontinuity in mixed audio sources. Because the phase of each sources are different, phases in the mixed sources are discontinued with overlap-add based TSM method.

The key idea of TSM algorithm is decomposing the audio signal in the short length with the analysis hop size H_a, and recomposing those frames with the synthesis hop size H_s. While $H_a < H_s$, the audio signal will be stretched after the procedure, and will be compressed if $H_a > H_s$.

In this section, we introduce some frequently used TSM algorithms and adventages and limitations of each methods.

2.1.1 Overlap-Add Method

Overlap-add method is the most simple and basic structure of TSM algorithms. In the overlap-add method, the audio signal x is decomposed with fixed analysis hop size H_a and fixed frame size N. The decomposed frame is x_m, while m is the order of the frame. After the decomposition, frames are resynthesized with the fixed synthesis hop size H_s and fixed frame size N. To preserve the gain of the audio signal, the Hann window is applied before synthesizing.

Because overlap-add method cannot connect the phase between frames, there are lots of artifacts in the source modified with overlap-add method.

2.1.2 Waveform-Similarity Overlap-Add Method

Waveform-Similarity Overlap-Add (WSOLA) algorithm is the improved version of overlap-add method. To reduce the artifacts, WSOLA selects the frame with the most similar frame with the previous frame. Therefore, phase discontinuity does not occur with WSOLA algorithm if the audio signal with single source. However, phase discontinuity still exists with the mixed audio signal with WSOLA, and there are also a transient degradation problem with WSOLA.

2.1.3 Phase Vocoder Method

TSM method with phase vocoder (PV-TSM) is a method to solve the phase discontinuity problem with mixed audio source. Basically, PV-TSM decomposes frames and synthesize frames same as overlap-add method. However, before synthesizing, frames are converted in to frequency domain with Fourier transform, and connect phases between frames for all frequencies in PV-TSM. Therefore, there are no

phase discontinuities in the mixed audio source with PV-TSM. However, PV-TSM is very weak for percussive transients comparing to harmonic signals because PV-TSM is concentrated on connecting harmonic information.

2.1.4 Pitch-Synchronous Overlap-Add Method

Pitch-Synchronous Overlap-Add (PSOLA) method is the improved version of overlap-add method. In PSOLA algorithm, the pitch information of the audio signal is needed to modify the audio signal. Based on the pitch information, PSOLA first finds the pitch mark, which means the local peak that appears for every fundamental period of the audio signal. After that, PSOLA decomposes the audio signal into frames based on the pitch marks. In the synthesizing part, PSOLA changes the number of frames to change the length. Also, the pitch of the audio signal can be changed when PSOLA adjust the distance between frames. Therefore, changing the pitch without using resampling is possible with PSOLA, and this is the reason to use PSOLA especially for human voices. However, because PSOLA always needs the pitch information, the high quality pitch tracking algorithm is essential to use PSOLA algorithm effectively.

2.1.5 Harmonic-Percussive Source Separation

Harmonic-percussive source separation (HPSS) is the method to separate the harmonic signal and the percussive signal in the single audio source. As mentioned above, PV-TSM is strong to preserve the phase in mixed audio signals, but is very weak for the percussive transients. To overcome this problem, many algorithms are proposed, and using HPSS to separate the signal and use different algorithm to the harmonic source and the percussive source is one of them. If PV-TSM is applied to the harmonic source, and overlap-add method is applied to the percussive source, it is possible to preserve percussive transients with no phase discontinuity.

2.2 Dynamic Time Warping

Dynamic time warping (DTW) algorithm is the algorithm that measures the temporal similarity between two signals. With the DTW algorithm, you can find the optimal path between the two signals that indicates the same portion in the two signals.

2.3 Pitch Tracking Algorithm

Pitch tracking algorithm is the algorithm that measures the pitch of the audio signal. There are two ways to measure the pitch, time-domain method and frequency-domain method. In this research, we used time-domain pitch tracking algorithm called YIN, which uses the autocorrelation method to measure the pitch.

Chapter 3. Related Works

3.1 Changing Musical Expressions with Extracting Features

There are some works in digital audio effects field about changing musical expressions of singing voices and musical instruments with extracting features. Some studies tried to manipulate musical features like vibrato [1], pitch, tempo [11, 3], and spectral envelope [10] for changing musical expressions. These studies are the most basic research to change musical expressions, but because they manipulate variables artificially to change musical expressions instead of using actual recorded examples, they are cumbersome and sometimes unnatural.

3.2 Transfer Musical Styles to Synthesized Sources

Also, there are some studies to extract styles from recorded examples to transfer musical styles to synthesized sources. [12, 13, 15] However, in this case, it requires the additional information such as lyrics and scores, and it does not transfer expressions from audio to audio directly.

3.3 Aligning the Source Signal and the Target Signal

Because this system directly transfers expressions from audio to audio, it is important to align the source signal and the target signal. In previous studies, some researchers tried audio-to-audio alignment to align audio and additional information such as lyrics and scores. [6, 14] Because this additional information contains onset data, the system does not have to align every frame by frame accurately. Also, there are some studies to align the temporal alignment of two voice signals, [4, 5] but in this case, they do not have to align every frame by frame because they used lyrical information to align them. However, in this system, we try to align two audio signals without any additional information.

Chapter 4. Proposed Architecture and Implementation

4.1 System Overview

Figure 4.1 illustrates the overall processing pipeline of the proposed system. It is composed of three modules that extract acoustic features from both voice signals and process the source.

The first module extracts the timing information from both signals, and align the tempo of the source signal. In this process, we use both musical feature and lyrical feature to measure the timing of the singing signal more accurate for every frame. After the feature extraction, the time-scale modification algorithm is applied to modify the source signal.

After the temporal alignment procedure, the system extracts the pitch information of both signals and align the pitch of the source signal. To extract the pitch, YIN algorithm is used in this system. After the feature extraction, the pitch synchronous overlap-add (PSOLA) algorithm is used to align the pitch of the signal without distortion in the formant.

At last, the dynamics alignment module works. In this step, the system extracts dynamics feature from both signals with envelope detector, and multiply the difference of both envelope signals to the source signal to align the dynamics.

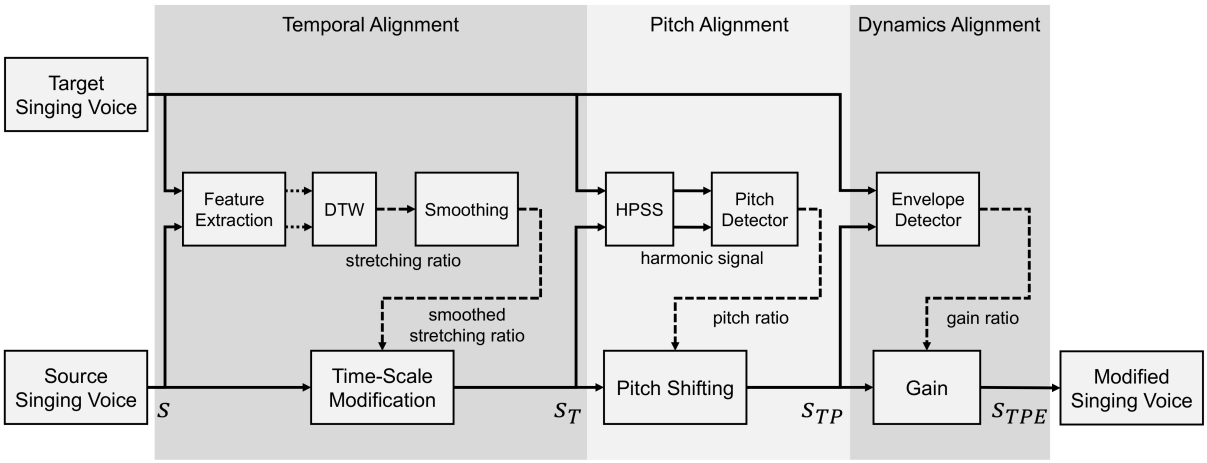

figure 4.1: System overview.

4.2 Temporal Alignment

The first step is temporal alignment that synchronizes note timings between the source voices. This is actually the most important step becasue the subsequent steps relies on the aligned source for pitch and dynamics processing. We basically use dynamic time warping (DTW), a dynamic programming algorithm which is popularly used for temporal alignment of music and audio data [18]. The issue here is what type of features will be used as input for DTW.

4.2.1 Feature Extraction

Considering that the source and target voices are rendered from the same song, one straightforward approach is transcribing the audio signals into MIDI notes and use the melody notes for DTW [17]. However, this approach can be affected by performance of the transcription module and, moreover, misses exploiting the phonetic information from lyrics which is another common part in the two singing voices. Thus, we instead extract audio features from the signals and use them for DTW.

Our initial approach was simply using the magnitude spectrum of two singings voices as audio features. However, the DTW algorithm often failed to find a correct alighnment path when either one voice has vibrato and pitch bending. The left in Figure4.2 shows the similarity matrix where each element was computed from cosine distance between every pair of the two magnitude spectra. The alignment path in red returned from the DTW algorithm tended to find the onset and offset of note quite successfully. However, it has severe detour, for example, that in the range of 300 to 350 time frames where the target voice has strong vibrato. This detour caused audible artifacts when the system modifies the time scale of the source signal.

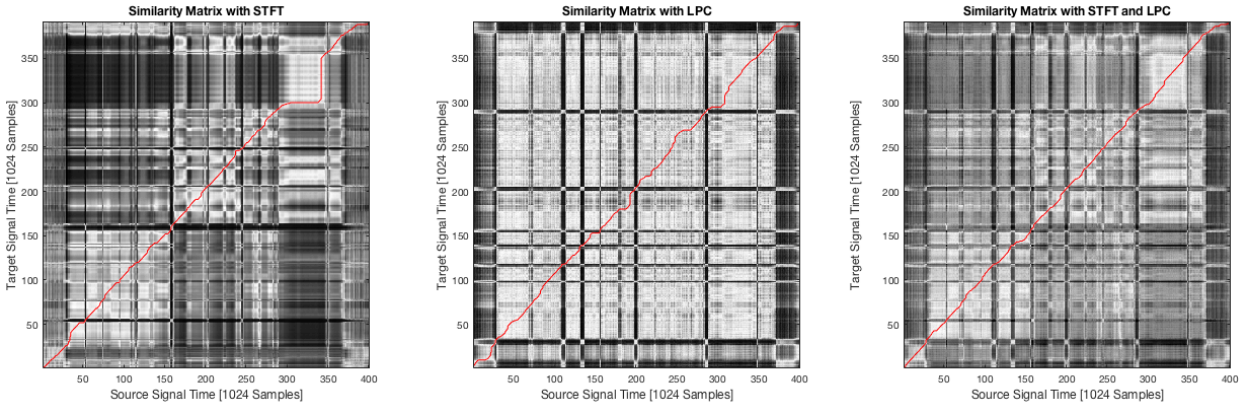

figure 4.2: *DTW path results with similarity matrices.*

We solved this problem using two methods. The first one is applying maximum filter to the spectrum. The maximum filtering is effective in suppressing vibrato or other pitch variations [8] and so this can help the detour problem. We used the maximum filter to the magnitude spectrum of both source and target before computing the similarity matrix as follows.

$$S_{max}(i,j) = \max(i, j-1:j+1) \tag{4.1}$$

where j corresponds to the frequency axis.

The second method is leveraging the phonetic information shared in lyrics of the song. The phonetic information tends to be less affected by musical expressions such as vibrato and pitch and so can allow more stable alignment. Since the phonetic information is related to the voice formant, we extracted the formant features using linear predictive coefficients (LPC). We chose the filter order according to [16]. Using the LPC, we compute a separate similarity matrix. The middle in Figure 4.2 shows the similarity matrix and alignment path by DTW. Compared to the DTW path by the spectrum, the detour in the segment with strong vibrato become more diagonal. When we listened to the processed sound, the path by LPC-only similarity matrix did not cause artifacts but often misses right note timings. Considering these advantages and disadvantage of both features, we create a new matrix by averaging

the two similarity matrices. The right in Figure 4.2 shows that it successfully reduces the detour problem and, at the same time, finds the accurate path.

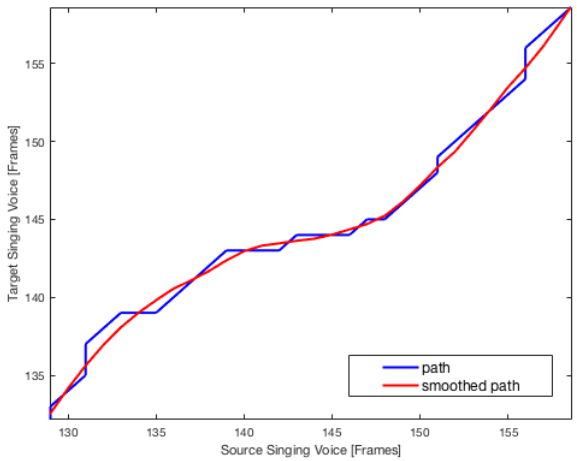

figure 4.3: *Raw path (blue) and filtered path with Savitzky-Golay filter (red).*

4.2.2 Smoothing Time Stretch Ratio

Given the alignment path, we need to find a sequence of stretching ratio to apply them for a time-stretching modification algorithm. Since the alignment path moves only three directions, upward, rightward and diagonal direction every frame, we need to smooth the path such that the stretching ratio is within a reasonable range.

...

To apply a TSM algorithm to the source signal, we changed the DTW path into an explicit function because an explicit function is easier to apply filters and calculate time-stretching rate. To convert DTW path to an explicit function, we removed the vertical interval in the path as follows.

Algorithm 1 Removing vertical interval in the DTW path

1: $expPath \leftarrow q(1), i \leftarrow 2$
2: **while** $i \leq length(p)$ **do**
3: **if** $p(i) \neq p(i-1)$ **then**
4: $expPath.append(q(i))$
5: $i \leftarrow i + 1$

To reduce the minor detours and to make the path smoother to reduce artifacts, we applied 3rd-order Savitzky-Golay filter to the path. The effect of Savitzky-Golay filter is shown in fig. 4.3.

To calculate the time-stretching rate α, the system simply used the slope of filtered path. Since one path value corresponds to one frame, we could apply the path slope to the time-stretching rate of each frame.

When correcting rhythm based on the path information, Time-Scale Modification(TSM) algorithm is used. In this system, we used TSM Toolbox, the open source MATLAB TSM algorithm code. [3] To obtain more sophisticated results, we used different algorithm to harmonic component and percussive component. A median filter was used to separate the harmonic and percussive components. After

the harmonic-percussive separation, phase vocoder was used to modify a harmonic component, and Waveform Similarity Overlap-Add(WSOLA) was used to modify a percussive component. The sum of modified harmonic component and modified percussive component is used for the rhythm corrected signal.

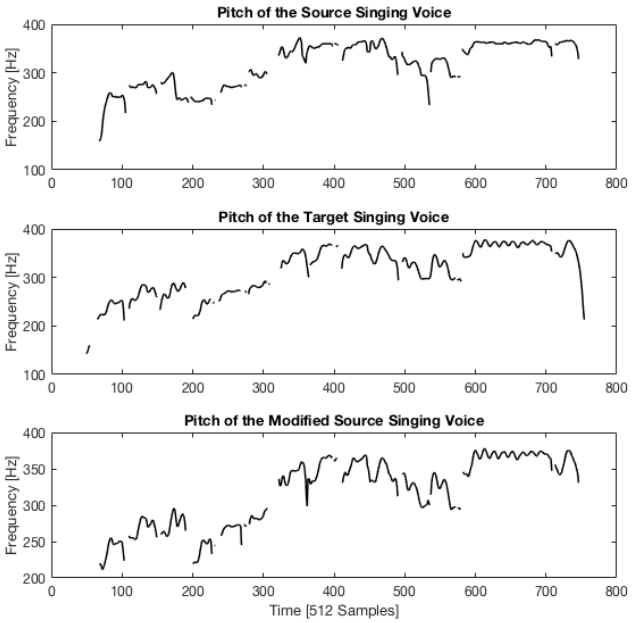

figure 4.4: *Pitch alignment.*

4.3 Pitch Alignment

To transfer the pitch of the target signal to the source signal, we used YIN algorithm to analyze the pitch of the signals. When we analyze the pitch, the unvoiced signal is excepted because it is difficult to measure, and is natural without changing the pitch.

To separate the unvoiced signal and voiced signal, the system uses aperiodicity of the signal. The part of the signal where the aperiodicity falls below 0.2 is regarded as an unvoiced signal and excluded from the pitch analysis and transplantation.

To reduce the unvoiced signal and get the stable pitch, the system uses harmonic-percussive source separation (HPSS) with median filtering [7] to separate the percussive signal and the harmonic signal from the singing voice. The system applies YIN algorithm to the harmonic signal to extract the more stable pitch.

Since the timing problem has already been solved in the rhythm phase, it is simple to calculate the pitch that needs to be changed based on the extracted pitch information. The beta value, the pitch amount that should be changed, is calculated as follows.

$$\beta(i) = \begin{cases} f0_t(i)/f0_{s_T}(i) & \text{if } aperiodicity > 0.2 \\ 1 & otherwise \end{cases} \tag{4.2}$$

$f0$ means the fundamental frequency of the signal, and $source*$ means the rhythm modified source signal.

The PSOLA algorithm is used to modify a pitch based on the extracted pitch information because it is an algorithm that can change the pitch without resampling. Since resampling causes the the formant break and changes the timbre of the voice, using PSOLA can retain the voice timbre of the signal.

4.4 Dynamics Alignment

The source signal, in which both the rhythm and pitch information are modified, is finally transplanted power of the target signal. The power of the signal is extracted envelope detector, which uses rms value. In this system, we use rms value to extract envelope instead of peak value because the envelope with peak value often outputs a negative number.

$$s_{TPE}[n] = s_{TP}[n] * env_t[n]/env_{s_{TP}}[n] \tag{4.3}$$

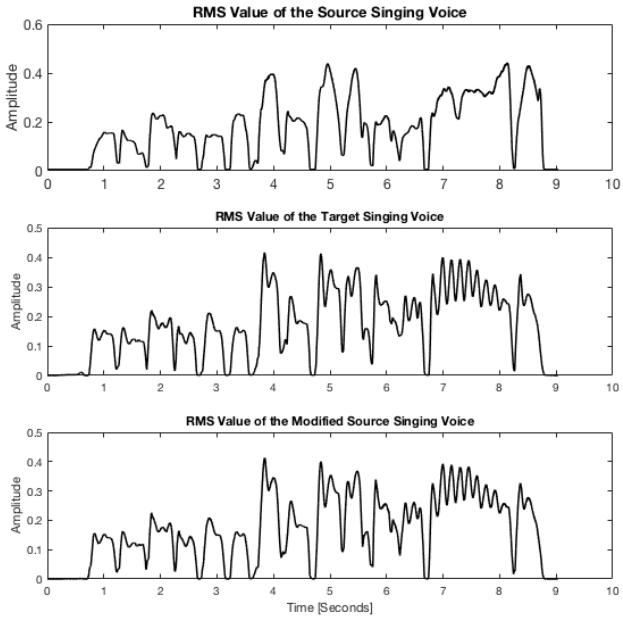

figure 4.5: *Energy alignment.*

Chapter 5. Evaluation

A quantitative and qualitative experiments are conducted to evaluate the performance of this vocal correction system. We used dynamic time warping to compare the path before and after aligning. Also, we asked people to evaluate the quality of converted signal.

5.1 Datasets

In this experiment, 4 songs were used as experimental data, and totally 12 modified signals were generated using one target signal and three source signals for each song. To verify that the system works well in various styles of songs, we chose the 4 songs with different styles. One of the four songs was the song for female vocals and the other three songs were male vocals. One of the three songs with male vocal was the song with swing rhythm, and other one was the song with low pitches.

5.2 Alignment Evaluation of the Converted Signal

To evaluate how well the alignment works, we used the DTW path of the similarity matrix made of STFT. If the slope of modified source's DTW path is more similar to diagonal line comparing to the slope of original source's DTW path, it means that the alignment works well.

Because the raw DTW path has only three kind of slopes (diagonal, vertical, horizontal), it is not appropriate to analyze the distribution of the slopes. To solve this problem, we used the average slope of 30 points in DTW path to get the slope distribution. To get the average slope, we used polynomial regression. In addition, to calculate the distributions of the larger and smaller slopes uniformly at 45 degrees, we used the arc tangent to express the slope as an angle.

Fig. **??** shows the variance of the DTW path slope of each source-target pair. The mean of variances of all DTW path slopes before the correction system was 0.06945, and it decreased to 0.0058492 after the source signal modified. Also, the variance of 11 of all 12 pairs decreased largely after modified.

5.3 Qualitative Evaluation

We used 12 pairs used in the previous evaluation to listen to the users and then conducted qualitative evaluation through surveys. The participants were asked to listen each pair and then evaluate how similar they were, how much vocal technique improved, and whether the modified signal was natural. For each question, the score ranged from 1 to 5.

Chapter 6. Conclusion

In this paper, we proposed the system that improves the vocal technique of the source signal through the alignment with the target signal. The system improves the technique of vocal by adjusting the alignment of the three elements of rhythm, pitch and energy using DTW, TSM algorithm and envelope detector, and it is possible to improve the vocal technique by using the result which shows the similarity with the target signal while maintaining the original vocal tone. However, the artifacts that occurs from the alignment process are expected to improve further.

Because these artifacts occur mainly in the time-stretching process of rhythm alignment, we plan to use onset detection with DTW to apply TSM algorithm more effectively.

Bibliography

[1] Axel Roebel, Simon Maller, and Javier Contreras, *Transforming vibrato extent in monophonic sounds*, Proc. of the 14th Int. Conference on Digital Audio Effects (DAFx), 2011.

[2] Jonathan Driedger and Meinard Müller, *A Review of Time-Scale Modification of Music Signals*, Applied Sciences, 6(2), 57, 2016.

[3] Jonathan Driedger and Meinard Müller, *TSM Toolbox: MATLAB Implementations of Time-Scale Modification Algorithms*, Proc. of the 17th Int. Conference on Digital Audio Effects (DAFx), 2014.

[4] Shimpei Aso, Takeshi Saitou, Masataka Goto, Katsutoshi Itoyama, Toru Takahashi, Kazunori Komatani , Tetsuya Ogata, and Hiroshi G. Okuno, *Speakbysinging: Converting singing voices to speaking voices while retaining voice timbre*, Proceedings of the 13th International Conference on Digital Audio Effects (DAFx), 2010.

[5] Takeshi Saitou, Masataka Goto, Masashi Unoki, and Masato Akagi, *Speech-to-singing synthesis: Converting speaking voices to singing voices by controlling acoustic features unique to singing voices*, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2007.

[6] Namunu C. Maddage and Khe Chai Sim and Haizhou Li, *Word level automatic alignment of music and lyrics using vocal synthesis*, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 6(3), 19, 2010.

[7] Jonathan Driedger, Meinard Muller, and Sebastian Ewert, *Improving time-scale modification of music signals using harmonic-percussive separation*, IEEE Signal Processing Letters, 21(1), 105-109, 2014.

[8] Sebastian Böck and Gerhard Widmer, *Maximum filter vibrato suppression for onset detection*, Proc. of the 16th Int. Conference on Digital Audio Effects (DAFx), 2013.

[9] Lindasalwa Muda, Mumtaj Begam, and Irraivan Elamvazuthi, *Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques*, arXiv preprint, 2010.

[10] Matthew Roddy and Jacqueline Walker, *A Method of Morphing Spectral Envelopes of the Singing Voice for Use with Backing Vocals*, Proc. of the 17th Int. Conference on Digital Audio Effects (DAFx), 2014.

[11] Nicholas J. Bryan, Jorge Herrera, and Ge Wang, *User-Guided Variable-Rate Time-Stretching Via Stiffness Control*, Proc. of the 15th Int. Conference on Digital Audio Effects (DAFx), 2012.

[12] Chih-Hong Yang, Pei-Ching Li, Alvin W. Y. Su, Li Su, and Yi-Hsuan Yang, *Automatic Violin Synthesis Using Expressive Musical Term Features*, Proc. of the 19th Int. Conference on Digital Audio Effects (DAFx), 2016.

[13] Pei-Ching Li, Li Su ,Yi-Hsuan Yang, and Alvin W. Y. Su, *Analysis of Expressive Musical Terms in Violin Using Score-Informed and Expression-Based Audio Features*, Proceedings of the International Symposium on Music Information Retrieval, 809-815, 2015.

[14] Simon Dixon, *Live tracking of musical performances using on-line time warping*, Proc. of the 8th Int. Conference on Digital Audio Effects (DAFx), 2005.

[15] Tomoyasu Nakano and Masataka Goto, *VocaListener: A singing-to-singing synthesis system based on iterative parameter estimation*, Proceedings of the Sound and Music Computing Conference, 343-348, 2009.

[16] Xuedong Huang, Alex Acero, and Hsiao-Wuen Hon, *Spoken Language Processing:A Guide to Theory, Algorithm and System Development*, 1st ed, Prentice Hall, 290, 2001.

[17] Roger B. Dannenberg, *An on-line algorithm for real-time accompaniment.*, International Computer Music Conference, Vol. 84, 1984.

[18] Meinard Müller, *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*, Springer, 2015.

[19] Jonathan Driedger, Meinard Muller, and Sebastian Ewert. *Improving time-scale modification of music signals using harmonic-percussive separation*, IEEE Signal Processing Letters 21.1, 105-109, 2014.

Acknowledgments in Korean

Curriculum Vitae in Korean

이 름: 용 상 언

학 력

2008. 3. – 2011. 2. 고양외국어고등학교

2011. 2. – 2015. 8. 한국과학기술원 전기및전자공학부 (학사)

2015. 9. – 2017. 8. 한국과학기술원 문화기술대학원 (석사)

학 회 활 동

1. **Sangeon Yong**, E.J. Lee, R. Peiris, L. Chan, and J. Nam, *ForceClicks: Enabling Efficient Button Interaction with Single Finger Touch.*, Proceedings of the Tenth International Conference on Tangible, Embedded, and Embodied Interaction. ACM, Yokohama (Japan), March., 2017.
2. E.J. Lee, **Sangeon Yong**, S. Choi, L. Chan, R. Peiris, and J. Nam, *Use the Force: Incorporating Touch Force Sensors into Mobile Music Interaction.*, Proceedings of the 13th International Symposium on Computer Music Multidisciplinary Research, Proto and Matosinhos (Portugal), September., 2017 (to be published).