v.morish'09: A Morphing-based Singing Design Interface for Vocal Melodies

Masanori Morise[1], Masato Onishi[2], Hideki Kawahara[3], and Haruhiro Katayose[4]

[1] College of Information and Acience, Ritsumeikan University,
1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577 Japan
[2] Graduate School of Systems Engineering, Wakayama University,
930 Sakaedani, Wakayama 640-8510, Japan
[3] Faculty of Systems Engineering, Wakayama University,
930 Sakaedani, Wakayama 640-8510, Japan
[4] School of Science and Technology, Kwansei Gakuin University,
2-1 Gakuen, Sannda, Hyogo, 669-1337 Japan
http://crestmuse.jp/index-e.html

Abstract. This paper describes a singing design method based on morphing, the design and development of an intuitive interface to assist morphing-based singing design. The proposed interface has a function for real-time morphing, based on simple operation with a mouse, and an editor to control the singing features in detail. The user is able to enhance singing voices efficiently by using these two functions. In this paper, we discuss the requirement for an interface to assist in morphing-based singing design, and develope an interface to fulfill the requirement.

Key words: Singing voice synthesis, voice morphing technique, user interface design

1 Introduction

Vocal manipulation is one of the most entertaining elements of computer music processing. Above all, replacing the singing style or voice characteristics of a singing voice with those of a professional singer has been a highly desirable function in vocal manipulation applications.

In the field of sound synthesis study, vocal synthesis has been a major research target, and core technologies have already been proposed [1],[2]. Commercial vocal synthesizers for PCs were released once computer and media technology had developed sufficiently, by 2005. In 2007, a vocal synthesizer, the conceptual basis of which is a virtual animation vocalist, *Hatsune Miku* (Vocaloid2), was released. As of late 2008, sales of Hatsune Miku had reached 40,000, a record-breaking number in desktop music software sales. Amateur creators are uploading their original songs, synthesized using Hatsune Miku, to video-sharing websites, and some of these videos have ranked in the top 10 most viewed. Users of vocal synthesis software edit lyrics and melody, i.e., pitch trajectory, using an editor that resembles a music sequencer. Some vocal synthesizing software provides

functions for automatic control of delicate pitch trajectories, such as portamento and vibrato, but users are obliged to elaborate on the pitch control parameters in order to obtain human-sounding singing voices. This operation is so troublesome that tools are being proposed for adjusting the parameters of vocal synthesizers so that the output resembles human singing voices [3].

We have been developing a morphing-based singing design technology capable of replacing the singing style or voice characteristics of a singing voice with those of a professional singer. In this paper, we discuss technologies and interfaces that allow the user to generate a newly synthesized voice by adjusting two morphing rates: one for pitch changes and one for voice characteristics.

The rest of this paper is organized as follows: Section 2 presents singing voice morphing and peripheral technology. Section 3 discusses requirements for morphing-based singing design and the development of *v.morish'09* to fulfill these requirements. Finally, Section 4 concludes this paper.

2 Singing design

In this section, we discuss a new parameter for use in singing design, singing voice morphing [4], and a peripheral technology, called STRAIGHT [5]. Generally, to facilitate understanding when we discuss singing voices, we use an example of a well-known singer instead of talking about the fundamental frequency (F0) or spectrum envelope. Although singing voices can be analyzed in terms of their F0 and spectrum envelope, it is difficult for us to explain the allure of the singing voice by citing these properties. Instead, we talk about the singing style and voice characteristics of a well-known singer. Therefore, although most vocal synthesizing software has an editor to control F0 and the spectrum envelope as the main parameters, it is difficult for creators to synthesize a desired singing voice by controlling these parameters. Synthesizing a natural singing voice might be simpler if the user could control a singer's style and voice characteristics instead of the F0 and spectrum envelope.

We propose a technique for singing voice morphing that controls singing style and voice characteristics by morphing the user's singing voices with a singer's characteristics. This technique enables the user to blend two features (singing style and voice characteristics) independently. The morphing of two singers' voices has also been proposed for Karaoke applications [1], but singing voice morphing differs from such applications in that there are two control parameters. A high-quality vocoder, STRAIGHT [5], is used for the singing voice analysis, morphing and synthesis. The user is able to model the singing voices by controlling the morphing rates of the singing style and voice characteristics. Furthermore, singing voice morphing also enables the user to control emotional color by blending emotional singing voices into the user's own singing voices.

The interface for STRAIGHT-based singing voice morphing consists of mapping the singing style and the voice characteristics to the horizontal and the vertical axes of a two-dimensional plane (Figure 1); the user is able to identify the morphing rates visually. This interface enables the user to reproduce

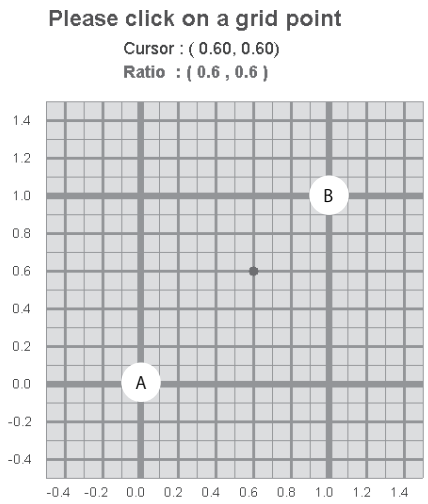

Fig. 1. GUI for singing voice morphing. The horizontal axis represents the morphing rate for the singing style. The vertical axis represents the morphing rate for the voice characteristics.

singing voices with morphing rates shown by the cursor. However, because STRAIGHT cannot achieve real-time analysis, morphing and synthesis, the user cannot change morphing rates during reproduction. In this paper, we use TANDEM-STRAIGHT[6] that produces the same results that STRAIGHT does, only much more quickly.

2.1 STRAIGHT and TANDEM-STRAIGHT

STRAIGHT, which is a vocoder system [7], analyzes a voice in terms of its fundamental frequency (F0), spectrum envelope and aperiodicity spectrogram (Figure 2). It is able to synthesize a voice that sounds as natural as a human voice captured by a microphone. In singing voice morphing, F0 is represented by the term "singing style." The spectrum envelope and aperiodicity are represented by "voice characteristics." General voice morphing [8] uses one control to morph these three parameters, and singing voice morphing has two parameters, corresponding to singing style and voice characteristics. STRAIGHT is incapable of real-time morphing and synthesis because they would require too much computational power. Therefore, the interface, shown in Figure 1, can only reproduce singing voices morphed in advance. The development of a real-time interface based on STRAIGHT has been difficult.

TANDEM-STRAIGHT [6] produces the same results that STRAIGHT does, only much more quickly. The problem of real-time synthesis is solved by using TANDEM-STRAIGHT. We have also developed a library in the C programming language, so that a real-time interface might be developed for TANDEM-

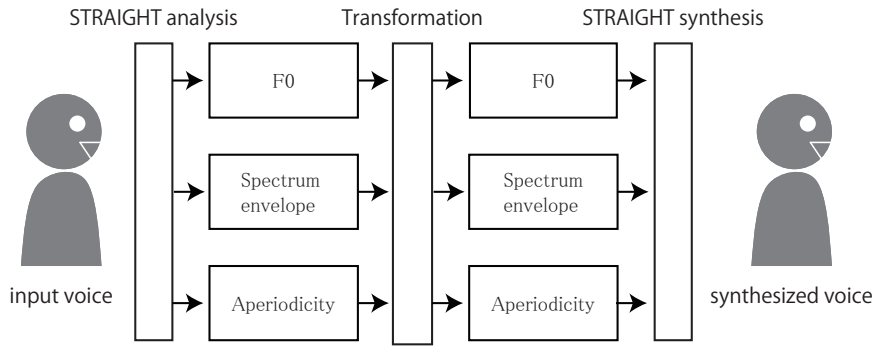

Fig. 2. Overview of STRAIGHT (TANDEM-STRAIGHT works the same way).

STRAIGHT. This library enables us to change morphing rates in real time during reproduction.

3 Design and development of the singing design interface

The interface supports morphing-based singing design and provides a method for aligning two singing voices. Most vocal synthesizing software is able to show the results of the user's input. Singing voice morphing must display the results of changing morphing rates so that the user will be able to perceive the difference in the quality caused by changing the morphing rate.

It is essential that the software include an off-line editor for editing singing voices. Previous software incorporates an off-line editor to control each parameter in singing design. Therefore, our interface will have an editor for controlling the morphing rates of singing style and voice characteristics. The user will be able to create the detailed results with this editor.

3.1 implementation of *v.morish'09*

We developed *v.morish'09* as an interface to fulfill the requirements described in section 3. Figure 3 is a screenshot of v.morish'09.

In Figure 3, the left-hand side of the interface provides real-time morphing control. The horizontal axis represents the singing style, and the vertical axis represents the voice characteristics (here, "voice color"). With this interface, the user can control morphing rates in real time during reproduction of the morphed singing voice, making it possible for the user to hear the changes being made. The right-hand side of Figure 3 shows an off-line editor for drawing the morphing rates of the singing style and the voice characteristics. The user is able to draw the morphing rate trajectory in detail using this off-line editor. Additionally, v.morish'09 can reproduce the morphed singing voices as modified by the morphing rate trajectory in the editor. Thus, the user can design singing voices by using these two functions.

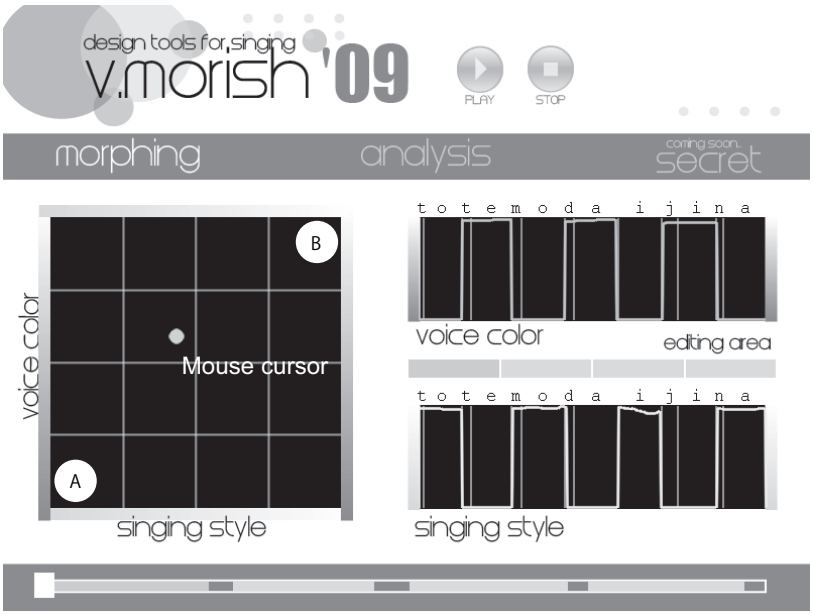

Fig. 3. Screenshot of the morphing-based singing design interface *v.morish'09*. This interface consists of two tools: a real-time GUI for morphing (left) and an editor for off-line processing (right).

3.2 Two procedures for the singing voice design

The user designs singing voices using the real-time interface both to draw a rough trajectory and to revise this trajectory.

Real-time control using the mouse The movement of the cursor during reproduction is reflected in real time by the morphed results. The change of the morphing rate is also shown in the trajectory represented within the editor. Therefore, detailed, real-time control is possible with this editor.

Detailed design using the editor The rough trajectory drawn in real time can be revised with the editor. The editor also shows phoneme boundaries, providing control over the morphing rates of each phoneme. Moreover, v.morish'09 is able to reproduce the morphed singing voices according to the morphing rate trajectory shown in the editor. The real-time interface cannot adjust the morphing rate trajectory in detail and reproduce by the same trajectory. Using the reproduction function, the user is able to reproduce the morphed singing voices according to the same trajectory.

3.3 Discussion

Many singing design methods and singing design interfaces have been proposed in the past. The proposed interface is different from them in control parameters that enable users to change singing style and voice characteristic directly. Replacing the singing style or voice characteristics of a singing voice with those of a professional singer will be possible by using v.morish'09. As v.morish'09 has the real-time control GUI and the editor to draw trajectory, the user may control these parameters easily. Evaluation of usability is important future work.

4 Concluding remark

In this paper, we discussed the requirements for morphing-based singing design, and proposed the interface v.morish'09 as a support tool of for morphing-based singing design. Our v.morish'09 enables users to control the singing style and voice characteristic directly. The evaluation of usability of v.morish'09 is important future work.

Acknowledgments This research was partly supported by the CrestMuse project, conducted by the Japan Science and Technology Agency (JST) and a grant-in-aid for young scientists (Start-up) 20800062.

References

1. Cano, P., Loscos, A., Bonada, J., de Boer, M., Serra, X.: Voice morphing system for impersonating in karaoke applications. In: Proc. ICMC. 109–112 (2000)
2. Bonada, J., Serra, X.: Synthesis of the singing voice by performance sampling and spectral models. IEEE Signal Processing Magazine, 24, 67–79 (2007)
3. Nakano, T., Goto, M.: VocaListener: An automatic parameter estimation system for singing synthesis by mimicking user's singing. In: Proc. IPSJ SIGMUS meeting, 49–56, May (in Japanese, 2008)
4. Kawahara, H. Ikoma, T., Morise, M., Takahashi, T., Toyoda, K., Katayose, H.: Proposal on a morphing-based singing design manipulation interface and its preliminary study. Journal of Information Processing Society, 48 3637–3648 (in Japanese, 2007)
5. Kawahara, H., Masuda-Katsuse, I., and de Cheveigné, A.: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction. Speech Communication, 27, 187–207 (1999)
6. Kawahara, H., Morise, M., Irino, T., Takahashi, T.: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In: Proc. ICASSP, 3933–3936 (2008)
7. Dudley, H.: Remaking speech. J. Acoust. Soc. Am., 11, 169–177 (1939)
8. H. Kawahara, R. Nisimura, T. Irino, M. Morise, T. Takahasni, and H. Banno, "Temporally variable multi-aspect auditory morphing enabling extrapolation without objective and perceptual breakdown," Proc. ICASSP 2009, pp.3905-3908 2009.