THE ROLE AND REALIZATION OF SOUND MORPHING AND HYBRIDIZATION IN COMPUTER MUSIC

Naotoshi Osaka
School of Science and Technology for future life,
Tokyo Denki University

ABSTRACT

Sound morphing and hybridization are new sound effects introduced by digital technology. First the definition of morphing is clarified, and traditional hybridization techniques are introduced, namely, the source-filter model and performance transcription. Moreover, new possibilities of hybridization are discussed using decomposition of harmonic structured sound to its self-similar harmonic structure. In this process morphing is also embedded, and the relationship among those is discussed.

1. INTRODUCTION

Morphing and sound hybridization can be achieved by digital technologies in recent years. Although these are classified to different technologies, they have things in common. One of those is that both are effects on a single stream.

Morphing is a technology developed in computer graphics which changing images from one to another continuously. It became popular by being used in movies and promotion videos in the early 90s.

The same analogy was taken to sound morphing: timbre is interpolated from one to another. It was also developed in the 90s.

Hybridization has been used since late 80s in music pieces. However, the definitions for those are not clear.

Clear definitions are not necessarily given to these technologies. Whatever definition a technology may have, an engineering interest of how the technology appeals to our intuition is realized with high quality and musical passages using those technologies are positioned in the overall music are of much importance.

In this paper, we first clarify the morphing problem, and introduce well known hybridization techniques. Then, a new concept using harmonic structure is introduced. This is a phenomenon which occurs at an intermediate unstable region between one stream and decomposed multi-stream state. Then the relationship between this type of hybridization and morphing is discussed.

2. SOUND MORPHING

We first clarify the morphing problem. Osaka proposed that sound morphing should be evaluated in three aspects[1]:
1.Continuity
2.Intermediateness
3.Homogeneity

In continuity, given two different sounds, timbre is continuously controllable. Timbre interpolation is another expression for this feature.

Intermediateness guarantees the interpolated timbre passes the point which is perceived as the central point between timbre of A and B.

In homogeneity, interpolated timbre should also keep the features that both sounds have in common. If a morphed sound satisfies both continuity and homogeneity, intermediateness is also maintained.

As an example, let us take a morphed sound between violin sound and cello sound. It should satisfy the timbre of bowed string, since it is the feature in common for both sounds. By satisfying this requirement and having other aspects being continuously changed, intermediateness is achieved.

Let us take more complex example: a morphed voice of a female voice to a male voice. We can count the natural human voice as a common feature. That is, a morphed sound should sound like a naturally spoken human voice. Therefore a robot voice or a degraded voice does not satisfy the requirement. This means that if two target sounds for morphing are natural sounds, that is, acoustic sounds, naturalness or sound quality is the important requirement for the morphed sound.

However, morphed sounds are not discussed in detail including item 2 and 3, and in general, only continuity is taken into consideration.

In the implementation, mostly sinusoidal model[2],[3] based technologies are well known in the early days[4],[5]. Nowadays straight-based synthesis gives a very high-quality speech morphing[6].

3. SOUND HYBRIDIZATION

On the other hand, sound hybridization is a synthesized sound which has timbre components from various sounds, satisfying a single sound stream. Two dominant methods are introduced as combining timbre components from various sounds in the next section.

Other possibility is as follows: given two sounds which melt together and becomes a single stream when mixed, a boundary area appears between a single stream and two streams. This narrow and unstable perceptive area can become a sound hybridization.

3.1. Adding phonemes to non speech using a source-filter model

Cross synthesis, a technology that hybridizes two sounds, was applied to music pieces from the early days. Of particular note, a synthesized sound adding the phonemes of a speech to noise is introduced in Joji Yuasa's piece in the late 80s[p1].

In its implementation, a source filter model is adopted. Speech is produced from a source of either expiratory flow noise or vocal fold excitation changed by dynamic behaviour of vocal tract caused by articulatory organs. A source-filter model simulates this process by multiplying a voice source and a vocal tract filter. While a speech is composed of a voice source and a vocal tract from the same subject, cross synthesis adopts these two components from two different sound streams.

Sound hybridization is not a sound mixture, but a synthesized sound of a single stream. Sound source and phoneme are different components of speech, and with this sound hybridization is conceptually established.

However, in reality, speech intelligibility is not good in many cases. The only successful case is when sufficient energy is supplied around formant frequencies for the source.

3.2. Adding musical expression

The aspects in common with the previous section and this section are, firstly that a stable sound source is selected which has an identity as a timbre and has some characteristics on the spectral domain. Secondly, from here we add higher-level components such as a phoneme filter and musical expressions.

Here, pitch events such as vibrato and power events such as rhythm are taken into consideration as musical expression. Adding these features to an instrument which is impossible to have such feature gives a hybridization.

In Osaka' s Morphing collage[p2], the sound of the sho, with which it is impossible to perform vibrato, is taken as a source and the vibrato of a shakuhachi performance is added. This hybridization enabled a sho to be performed with shakuhachi vibrato. The sho sound was synthesized by a physical model[7][8].

4. SOME OTHER HYBRIDIZATION METHOD

Some other methods besides adding extra information to the source are considered in this section. We take two homogeneous sounds, two sounds having harmonic structure as an example. We can think of two statuses: one is these two are unified and function as a single stream, and the other is completely separated and two streams are there. If there is an intermediate state in between them, sound morphing can be achieved by continuously changing the state. This intermediate state can be a sound hybridization, since it is not a double stream and yet you can feel two aspects of sounds.

4.1. Sound hybrid in harmonic structure

As the simplest example, a sound of harmonic structure is taken. Let sound A be a harmonic structure with fundamental frequency f0. A can be divided into two series of other harmonic structure: B odd number multiple of f0, and C even number multiple of f0. This means that sound stream A is decomposed to sound stream B and C with fundamental frequency f0 and 2*f0 respectively. This process is repeated to the stream C, and self similarity is imbedded in harmonic structure. However, this structure gives octave relation, and it is not always musically interesting. This is shown in Fig. 1.

In the example above, each component does not overlap and perfectly decomposed to other harmonic structured sound stream. One of the other examples which do not allow overlap is the odd numbered harmonic structure and six multiple harmonic structure. This gives a perfect 12th interval. If we allow overlap of some components, odd numbered harmonic and three multiple harmonic structure give two streams of the perfect 5th interval. This is shown in Fig. 2.

As seen above, if overlap is allowed, harmonic structure is regarded as a sound stream equipped with a self-similarity structure. And if we can achieve an intermediate state between one unified structure and multiple structure, this can become a sound hybridization. Fig. 3 depicts a diagram of such hybridization.

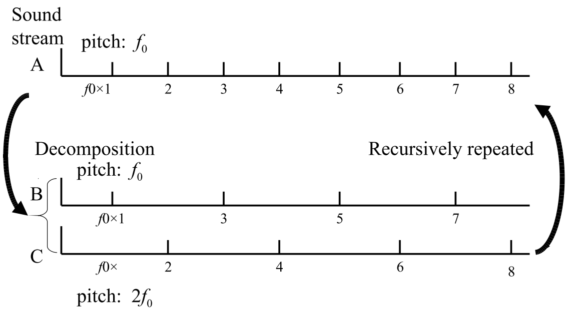

Figure 1. Decomposition of harmonics (1)

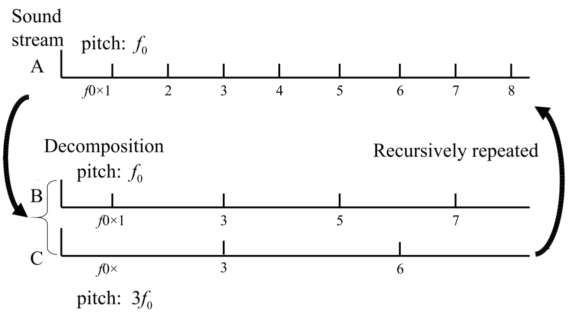

Figure 2. Decomposition of harmonics (2)

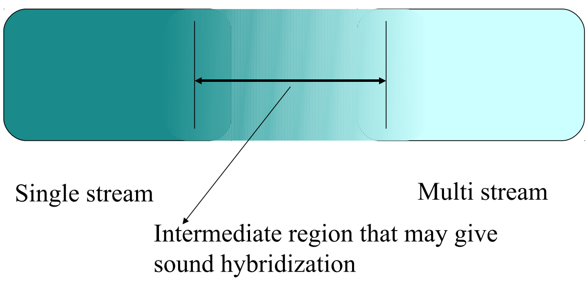

Figure 3. Boundary region of a single stream and multi stream

4.2. Possibility of generating sound hybridization based on harmonic structure

It is difficult to construct a sound hybridization from a sound stream which is composed by a pure tone and its harmonic structure, since all the possible sound streams embedded within are unified, and a sound with the same fundamental frequency of the pure tone is perceived. However, by changing some conditions an intermediate state can be expected. Here are some examples.

1. Emphasis of power of the specific harmonics Harmonics can be heard separately by attention. As seen in Mongolian singing, Khoomii, the emphasis of specific harmonics can decompose a single stream into two.
2. The f0s of two candidates of streams are slightly moved from the integer multiple relation It is known that harmonics with slightly different frequency from the integer multiple frequency of f0 still stands as a single stream. Therefore this gives room for having an intermediate state by changing a complete inter-multiple frequency of f0 to gradually apart from that relation.
3. Usage of non pure-tone based harmonic structure but pseudo harmonic structure based sound such as natural acoustic sounds. Natural voiced sound such as ordinary speech has a harmonic structure, but it is not completely an integer multiple of f0, and each harmonic has amplitude or frequency modulation. Some harmonics are damaged and noise is supplied in the frequency in question. This gives an intermediate state as well.
4. Sound generation in room acoustics By assigning a decomposed embedded harmonic structure to each channel of stereophonic speakers, respectively, we can expect an intermediate state. Supposing A is decomposed into B and C as in 4.1. If written in terms of sound level, from left channel, a*B+(1-a)*C sound is output, while (1-a)*B+a*C is output from the right channel, where a represents a fader level as a weight. By changing the a, that is, panning level, we can hear single monaural sound A to decomposed sound B from the left and C from the right speaker. We can expect the in-between states in some point in a's range.

4.3. The similarity and different points of morphing and hybridization

In morphing, there is a way to present intermediate timbre steadily. However, most applications adopt the morphing changing dynamically from a sound to another, and the presenting time of an intermediate timbre is relatively short. This is because in spite of stable sound A and B, the intermediate timbre is unstable and moving quickly from one sound to another is an essential factor of the morphing. By unstable we mean the perceptive reality of the sound is not clear and is not perceived as one particular image. Therefore by changing the timbre in a short time, an aspect of sound illusion can be realized.

Sound hybridization as well, it is hard to recognize two particular components as a single stream from two different sounds, and unstable region where either a single stream or double streams can be heard back and forth. Therefore, the morphed sound which includes hybridized sound in between two sounds can be more effective, rather than presenting a steady state hybridized sound.

4.4. Application to a music piece

The piece gA dialogue with Maxh is one which is performed in concert III of the ACMP concert series [p3]. Cross synthesis was done for Max Mathews' voice as a phoneme filter, and musical sound as a source. When it was premiered, composer' s voice speaking nonsense Japanese words was used. The revised version this time, uses the recorded voice of his lecture. Phoneme filter is made from the fft cepstrum analysis and used up to 100th order. Using spectral envelope acquired from these cepstra, sound hybridization with source sounds based on bells and mixed sounds of those was synthesized.

5. CONCLUSIONS

Definitions of sound morphing and sound hybridization are given and application examples are introduced. Some classification of those concepts are introduced. One way is to add higher-level information such as phoneme or performance expression to a stationary timbre. Moreover, as a different type, methods based on harmonic sound decomposition were introduced. A harmonic structure has another harmonic structure in itself and has a multiple sound streams as self-similar structure. Two states are A: all the harmonic structure are assembled as one. And B: multiple harmonic structures exist after decomposition. In the boundary of these unification and decomposition, intermediate regions can exist. Sound hybridization was discussed in those regions and the possibility of future direction of the field was considered. Although the former style was seen frequently, various unstable intermediateness based hybridizations are a new field and should be explored.

6. REFERENCES

[1] Naotoshi Osaka, "Timbre morphing and interpolation based on a sinusoidal model," Proc. of ICA/ASA joint meeting, pp.83-84, Seattle, June 1998.

[2] McAulay, R. J. and Quatieru, T. F., "Speech Analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, & Signal Process.*, Vol. ASSP-34, No. 4., pp.744-754, Aug. 1986.

[3] Serra, X. and Smith, J. "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System based on a Deterministic plus Stochastic Decomposition," Computer Music Journal, MIT Press, 14(4), pp. 12-24, 1990.

[4] Tellman E., Haken L., and Holloway B, "Timbre morphing using lemur representation," Proc of ICMC '94, pp.329-330, 1994.

[5] Naotoshi Osaaka, "Timbre interpolation of sounds using a sinusoid al model," ICMC 95 Proceeding, pp. 408-411, Banff, Canada, Sep. 1995.

[6] Kawahara, H.; Nisimura, R.; Irino, T.; Morise, M.; Takahashi, T.; Banno, H.; , "Temporally variable multi-aspect auditory morphing enabling extrapolation without objective and perceptual breakdown," Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on , vol., no., pp.3905-3908, 19-24 April 2009

[7] Takafumi Hikichi, Naotoshi Osaka, and Fumitada Itakura, "Time-domain simulation of sound production of the sho," The Journal of the Acoustical Society of America, Vol. 113, No. 2, pp. 1092-1101, February 2003.

[8] Takafumi Hikichi, Naotoshi Osaka, and Fumitada Itakura, "Sho-So-In: Control of a physical model of the sho by means of automatic feature extraction from real sounds," Journal of New Music Research, Vol. 33, No.4, pp.355-365, 2004.

7. PERFORMANCES

[p1] Joji Yuasa, "*The nine levels by Ze-ami* for seenteen players and computer generated quadraphonic tape," commissioned by IRCAM (1988), NEUMA records (1997)

[p2] Naotoshi Osaka, "*Morphng Collage* for piano and computer," organized by Ensemble Vivo, Tokyo Opera City, Dec 19, 2002.

[p3] Naotoshi Osaka, "A dialogue with Max for violin and electroacoustics," ACMP 2011, Tokyo. 2011.