Journal of New Music Research
2004, Vol. 33, No. 4, pp. 355–365

Sho-So-In: Control of a Physical Model of the Sho by Means of Automatic Feature Extraction from Real Sounds

Takafumi Hikichi[1], Naotoshi Osaka[2] and Fumitada Itakura[3]

[1]NTT Communication Science Laboratories, NTT Corporation, 2–4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan; [2]School of Engineering, Tokyo Denki University, 2–2 Kanda-nishiki-cho, Chiyoda-ku, Tokyo 101-8457, Japan; [3]Faculty of Science and Technology, Meijo University, 1–51 Shiogamaguchi, Tenpaku-ku, Nagoya, Aichi 468-8502, Japan

Abstract

This paper proposes a synthesis framework for sound hybridization that creates sho-like sounds with articulations that are the same as that of a given input signal. This approach has three components: acoustic feature extraction, physical parameter estimation, and waveform synthesis. During acoustic feature extraction, the amplitude and fundamental frequency of the input signal are extracted, and in the parameter estimation stage these values are converted to control parameters for the physical model. Then, using these control parameters, a sound waveform is calculated during the synthesis stage. Based on the proposed method, a mapping function between acoustical parameters and physical parameters was determined using recorded sho sounds. Then, sounds with various articulations were synthesized using several kinds of instrumental tones. As a result, sounds with natural frequency and amplitude variations such as vibrato and portamento were created. The proposed method was used in music composition and proved to be effective.

1. Introduction

This paper proposes a sound synthesis technology that produces rich and expressive timbres for music composition and content creation. A physical model of a sho is used to obtain sonorities that cannot be realized by musical instruments. Our previous paper has shown that the proposed physical sho model has similar physical characteristics to an actual instrument (Hikichi et al., 2003). This paper concentrates on the control issues that relate to creating rich and expressive timbres using this model.

A sho is an Asian free-reed instrument, and this family of instruments has spread from east to south Asia. The sho is composed of a cavity part with a mouthpiece and seventeen bamboo pipes with finger holes, and metal reeds are glued to the lower side of the pipes inside the cavity (Figure 1). In Japan, the sho is used to play chords or tone clusters in traditional gagaku music. Although many attempts have been made to produce more articulatory and dynamic sounds in contemporary music, there are limitations that arise from the sho's structure. For example, it is difficult to play notes with large pitch changes such as portamento.

One of the merits of the use of physical models in the computer music context is that, unlike real instruments, the values of the parameters of the model can be modified without any loss of their timbral identity, and hence the model can be used to explore timbre (Burtner & Serafin, 2002; Roads, 1996; Smith, 1996). In this study, we use this flexibility to implement articulations that we can find in other musical instruments, and attempt to extend the sho timbre space.

This research was carried out to develop sound hybridization techniques and is an extension of the notion of cross-synthesis (Mathews et al., 1961; Moorer, 1979; Tellman, 1994). Cross-synthesis combines two sounds, such as a human voice and an orchestra, to produce a single composite sound. Here, a synthesis model replaces one sound, and the other sound is used to extract articulations.

In general, it tends to be difficult to estimate the physical parameters of the model correctly from a given acoustical signal (Hélie et al., 1999). D'haes and Rodet (2003) have tackled this problem for the trumpet using two perceptual

Correspondence: Takafumi Hikichi, NTT Communication Science Laboratories, NTT Corporation, 2–4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan. E-mail: hikichi@cslab.kecl.ntt.co.jp

DOI: 10.1080/0929821052000343831 © 2004 Taylor & Francis Group Ltd.

Fig. 1. A sho.

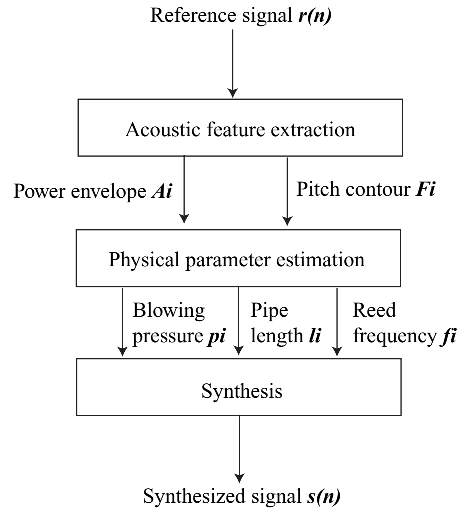

Fig. 2. Block diagram of Sho-So-In system.

similarity criteria related to spectral envelopes and fundamental frequency. This paper describes a new synthesis algorithm that can synthesize sounds with the same articulations and ornaments as the reference signal, namely the acoustic input.

2. Sho-So-In: synthesis system based on a physical model of the sho

This section describes the configuration of Sho-So-In, our sound synthesis system. Sho-So-In is an abbreviation of "<u>Sho</u> <u>So</u>unds <u>In</u>teresting".[1] The main features of the system are as follows. First, it is expected to create natural sounds because the synthesis is performed by a physics-based method. That is, by specifying proper physical parameters, the model calculates sound samples automatically according to physical law. Second, real musical tones are used to extract the control parameters, and this makes it possible to achieve precise control and specification. Thus, users can specify their desired articulations by using acoustic signals that have such characteristics.

Sho-So-In consists of the following three components:

- Acoustic feature extraction
- Physical parameter estimation
- Synthesis

The system configuration is shown in Figure 2. Each component of the system is described below.

2.1 Acoustic feature extraction

When the input signal (referred to as the reference signal) is given, this system tries to produce a sho-like sound with the articulation of the input signal. Here, there are many acoustic features that can be considered "articulation". In this study, the fundamental frequencies and power per frame are used as the most fundamental acoustic features. We will refer to these time series data of the fundamental frequencies as the pitch contour, and refer to the time series of the frame power as the power envelope. We used a cepstrum-based method to extract the pitch contour (Noll, 1967).

2.2 Physical parameter estimation

2.2.1 Selection of control parameters

In a previous study (Hikichi et al., 2003), we presented physical parameters to simulate one tube (B4) and compared the simulation with measured results. Typically, a sho has more than 15 sounding pipes, and the appropriate parameter values for each pipe are different. However, because we intend to use the model as a synthesis tool, it is desirable to be able to control it with a small number of parameters. Hence, a preliminary investigation was undertaken and the following three dominant parameters were selected (Hikichi et al., 2002).

- Blowing pressure p
- Pipe length l
- Mode frequency of the reed f

These parameters are used as control parameters. Of these parameters, only blowing pressure p can be controlled in the case of a real instrument. A player changes the acoustical length of the pipe by opening and closing finger holes, but this action only provides on/off control by changing the oscillation condition. Changing length l of the model provides more precise timbre control, and also pitch control.

2.2.2 Determination of pipe length l *and reed frequency* f

This section describes how to determine parameters l and f when the pitch contour is given.

The fundamental frequency of the synthesized tone mainly depends on l and f. The effect of p on the fundamen-

[1] Sho-So-In is also the name of a repository at Todaiji temple where ancient treasures including musical instruments have been preserved for more than a thousand years.

tal frequency is negligible. For example, a 5% increase in p from 800 to 840 Pa caused less than a 0.1-Hz increase in the fundamental frequency. In contrast, the same 5% increase in pipe length l and reed frequency f caused a 2.5-Hz decrease and a 20.8-Hz increase, respectively. Hence, sounds were synthesized using different l and f pairs with constant p, and analyzed to obtain the pitch. We refer to this correspondence between pitch and (l, f) pair as the pitch table. We have already shown in Hikichi et al. (2003) that the oscillation condition is satisfied when the resonance frequency of the pipe is lower than the reed frequency. Furthermore, the pipe resonance frequency is about three-fourths of the reed frequency in a real instrument. Based on this knowledge, we selected (l, f) pairs and undertook the synthesis. Using the pitch table obtained above, we determined parameters (l, f) using the following procedure.

1. Assume the fundamental frequency of a reference signal for frame i to be F_i, and search for the nearest frequency $\hat{F}_i$, and the second nearest frequency $\bar{F}_i$ from the pitch table. Here, $\hat{F}_i$ and $\bar{F}_i$ should be on either side of F_i.
2. Assume that the parameters corresponding to the fundamental frequencies $\hat{F}_i$ and $\bar{F}_i$ are $\hat{P}$ and $\bar{P}$, respectively. The parameter for the i-th frame is calculated by the interpolation of $\hat{P}$ and $\bar{P}$ at a ratio of distances from F_i to $\hat{F}_i$ and from F_i to $\bar{F}_i$.
3. Repeat procedures 1 and 2 for each frame.

2.2.3 Determination of blowing pressure p

Our preliminary investigation showed that the power envelope of the synthesized sound is a monotonically increasing function of the blowing pressure parameter. However, to cause the oscillation to occur requires a certain amount of pressure exceeding the threshold pressure. So, the mapping between the power envelope of the reference and the blowing pressure should be nonlinear such that a small increase in the power envelope in the low range corresponds to a large increase in the blowing pressure. Hence, we assume the n-th root as the nonlinear mapping function from frame power to blowing pressure, and we determine the optimal n value experimentally.

That is, if we assume the power envelope of the reference signal is A_i, the blowing pressure is calculated by $p_i = P_b\sqrt[n]{A_i}$. Here, $\sqrt[n]{A_i}$ is normalized by its maximum value, and P_b is the maximum blowing pressure needed to adjust the blowing pressure to the proper value for synthesis of the normal sound. P_b = 800, 1200, 1600 Pa is used based on the observation of a real instrument playing.

According to the procedures described in 2.2.2 and 2.2.3, the control parameters for synthesis (p_i, l_i, f_i) are specified for each frame time.

2.3 Synthesis

Synthesis is undertaken using our physical model of the sho. At the synthesis stage, control parameters are interpolated with time. Our physical model of the sho is briefly described. A more detailed derivation can be found in Hikichi et al. (2003).

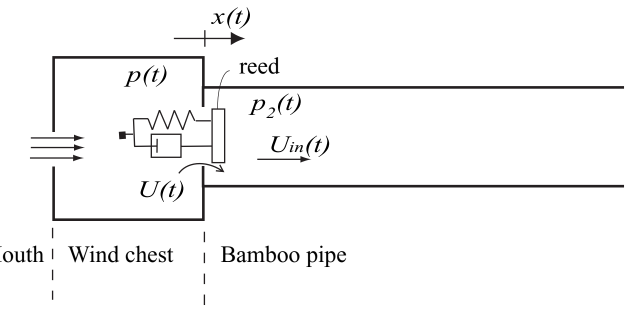

Fig. 3. Physical model of the sho.

The basic physical model is described by the following equations:

$$\frac{d^2x}{dt^2} + \frac{\omega_r}{Q}\frac{dx}{dt} + \omega_r^2 x = \frac{1.5\,WL}{m}(p(t) - p_2(t)), \tag{1}$$

$$p(t) = p_2(t) + \frac{\rho}{2}\left[\frac{U(t)}{CF(x)}\right]^2 + \frac{\partial}{\partial t}\left[\frac{\rho U(t)\delta}{CF(x)}\right], \tag{2}$$

$$F(x) = W\left[x^2 + b^2\right]^{\frac{1}{2}} + 2L\left[0.16x^2 + b^2\right]^{\frac{1}{2}}, \tag{3}$$

$$p_2(t) = Z_0 U_{in}(t) + r(t) * (p_2(t) + Z_0 U_{in}(t)), \tag{4}$$

$$U_{in}(t) = U(t) + 0.4WL\frac{dx}{dt}, \tag{5}$$

$$r(t) = -\alpha \exp\left\{-\beta(t - 2l/c)^2\right\}. \tag{6}$$

Equation (1) describes the motion of the reed when pressure p is applied inside a wind chest (Tarnopolsky et al., 2000), where p_2 is the pressure just under the reed, x is the displacement at the tip of the reed, Q is the resonance Q value, and ω_r is the angular frequency. W, L, and m are the width, length, and mass of the reed, respectively.

Nonlinear coupling between the reed and the pipe is described by Bernoulli's equation, i.e., Equations (2) and (3), where $U(t)$ is the volume velocity through the slit, $F(x)$ the area of the slit, C the flow contraction coefficient, ρ the air density, δ the inertia parameter, and b the clearance gap around the reed.

Equations (4)–(6) are employed to calculate the pressure at the entrance of the tube p_2, where Z is the characteristic impedance of the tube, $U_{in}(t)$ is the net volume velocity input into the tube, $r(t)$ the reflection function, and the asterisk denotes convolution.

By discretizing Equations (1)–(6), pressure p_2 and volume velocity $U_{in}(t)$ can be calculated recursively.

Radiated sound pressure is calculated using the transfer function of a pipe. The transfer function from the volume velocity at one end of a pipe and the pressure at the other end can be calculated assuming the shape and boundary condition of the pipe (Caussé et al., 1984). This method has also been used in previous studies, such as (Adachi & Sato, 1995), for modeling a brass instrument. In Adachi & Sato (1995), radiation loss was calculated on the assumption that the spherical wave radiates from the bell of the instrument. Here we also assume spherical radiation at the boundary condition. Using this transfer function, the radiated pressure is calculated from the volume velocity obtained by Equations (1)–(6).

3. Experiment

3.1 Evaluation criteria

Two kinds of objective criteria are used to evaluate how well the articulation of the reference signal is conveyed to the synthesis signal.

3.1.1 Power correctness

The difference between the power envelopes of the reference and the synthesized sound is expressed by the signal to deviation ratio (SDR).

$$SDR[dB] = 10\log\left(\frac{\sum_{i=0}^{N-1} A_i^r}{\sum_{i=0}^{N-1}\left|A_i^r - A_i^s\right|}\right).$$

A_i^r: normalized power envelope of the reference, A_i^s: normalized power envelope of the synthesized sound, where i denotes frame number.

3.1.2 Pitch correctness

Pitch correctness is defined as the ratio of the number of frames whose error is less than 5 Hz to the total number of frames.

3.2 Reproduction of articulations using sho sounds

3.2.1 Experimental conditions

During the acoustic parameter extraction, the acoustic feature is extracted using a 20 ms window and a 5 ms shift, and pitch extraction and voiced/unvoiced discrimination based on the cepstrum method are undertaken (Noll, 1967). During the physical parameter estimation, the acoustic parameters are converted to time series data of the physical parameters, and synthesis is performed. The pitch contour and power envelope are then extracted from the synthesized sounds in the same manner, and compared using the criteria described in Section 3.1. In order to create a pitch table, 30 pairs of (l, f) values with constant p values were used to synthesize sounds as described in Section 2.2.2. A pitch table for fundamental frequencies of 400–535 Hz was obtained.

3.2.2 Preliminary investigations of the system

First, synthesized sho sounds were used as a reference signal.

Synthesized signals with a constant amplitude that were used to create the pitch table were input as a reference, and synthesis was performed. As a result, we obtained a pitch correctness of 99.7%. Then, a pitch contour that changed linearly with time was provided manually, and the physical parameter was estimated and synthesis performed. In this case, almost the same level of performance was obtained.

These results show the effectiveness of the parameter estimation. Although there is a slight discrepancy between the reference and synthesized sounds, it is concluded that parameter estimation works very well with the static signals used here.

3.2.3 Performance for recorded sho sounds

Next, natural recorded musical sounds are applied to the system. To begin with, recorded sho sounds are used as a reference signal, and the power envelope and pitch contour are compared. The n value and maximum blowing pressure P_b of the mapping function $p = P_b\sqrt[n]{A}$ are used as parameters.

The recorded sho sounds used here are naturally blown tones with no specific articulations. Their amplitude gradually increased, and decreased, and the duration was about 10 s. Two samples with pitches A4 and B4 were analyzed. It should be noted that a sho is tuned slightly lower than modern Western musical instruments.

Figure 4 shows the power envelope correctness. It shows a peak when $n = 4$ or 5. The pitch correctness was about 80% as shown in Figure 5. An informal listening test showed that the n value should not be made too large, because it would also make sounds in the silent parts such as at the beginning. In this part, pitch estimation might fail in the analysis stage, and this would lead to improper perceptual effects. Hence, $n = 4$ is used hereafter. As for P_b, there were no big differences both in the power envelope correctness and in the pitch correctness among P_b = 800, 1200, and 1600 Pa. Hence, $P_b = 800$ was used hereafter.

We then undertook further detailed investigations. The pitch contour and power envelope are plotted in Figure 6. With the synthesized tone, it was found that the pitch tends to rise slightly with increases in blowing pressure, which is different from the case of the recorded tone. The power envelope shows a nice correspondence.

We found that the pitch contour was not extracted reliably at the beginning and ending parts because of its small amplitude. This effect is included in the pitch correctness measure described in 3.1, and this may introduce error into the measure. To avoid this kind of error, hereafter we consider only frames where both the reference and the target are judged to be voiced. This modified measure exceeds 99%.

Next, we used tonguing articulation as an example of more dynamic sounds. This articulation is not commonly

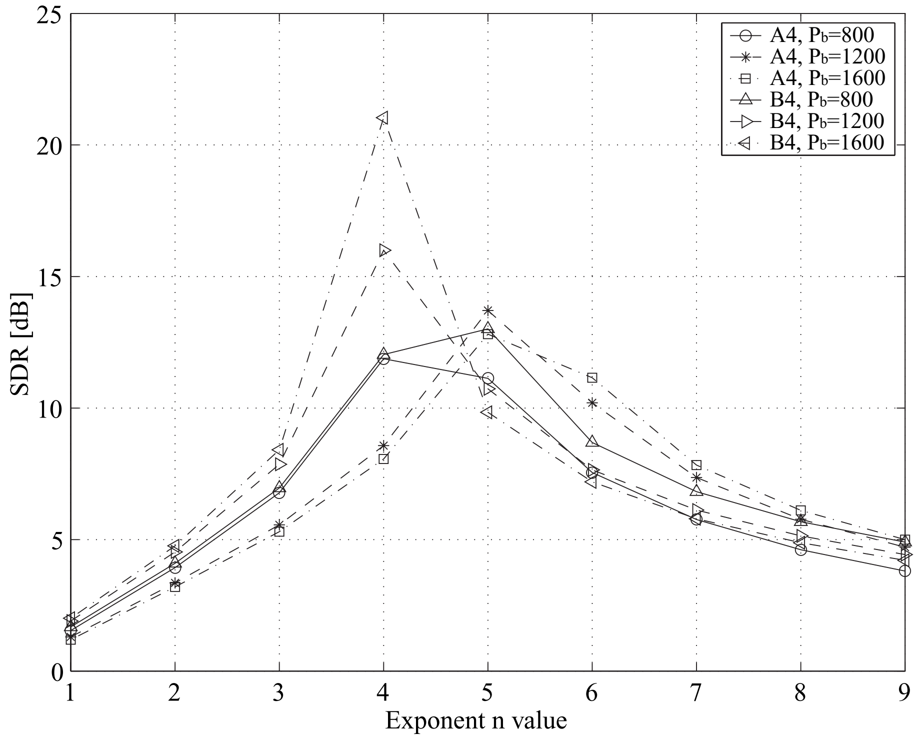

Fig. 4. Correctness of power envelope vs. exponent n value.

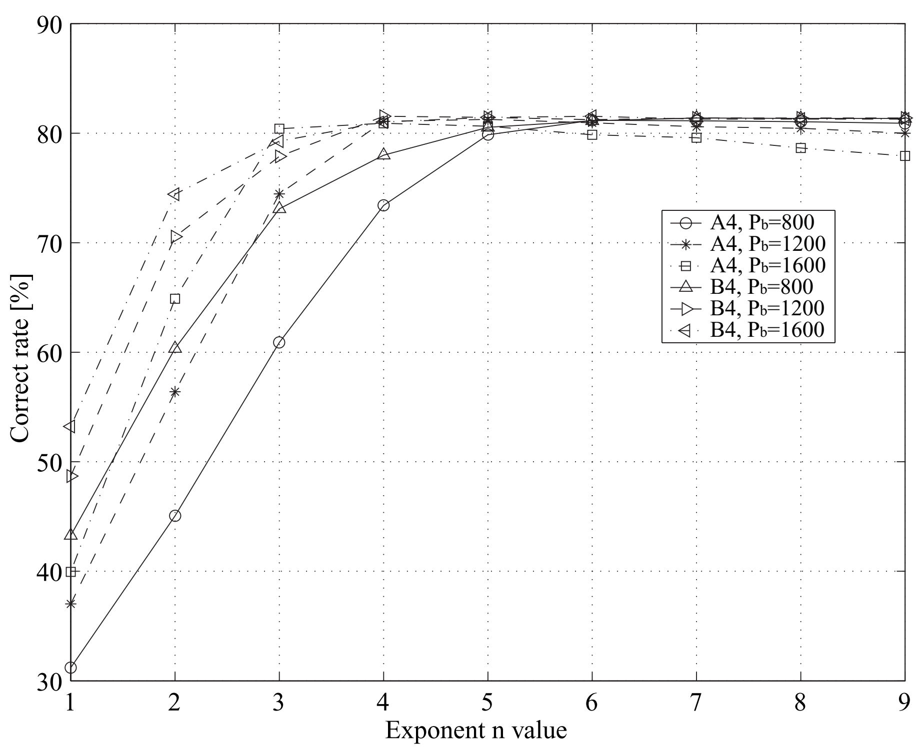

Fig. 5. Pitch correctness vs. exponent n value.

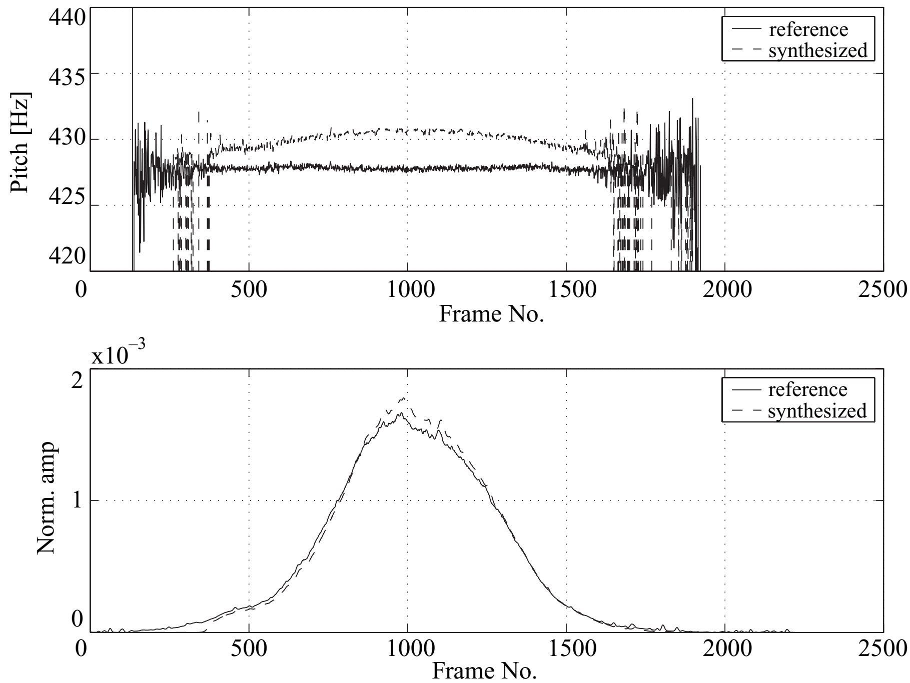

Fig. 6. Pitch contour and power envelope extracted from recorded and synthesized sounds (A4, normal).

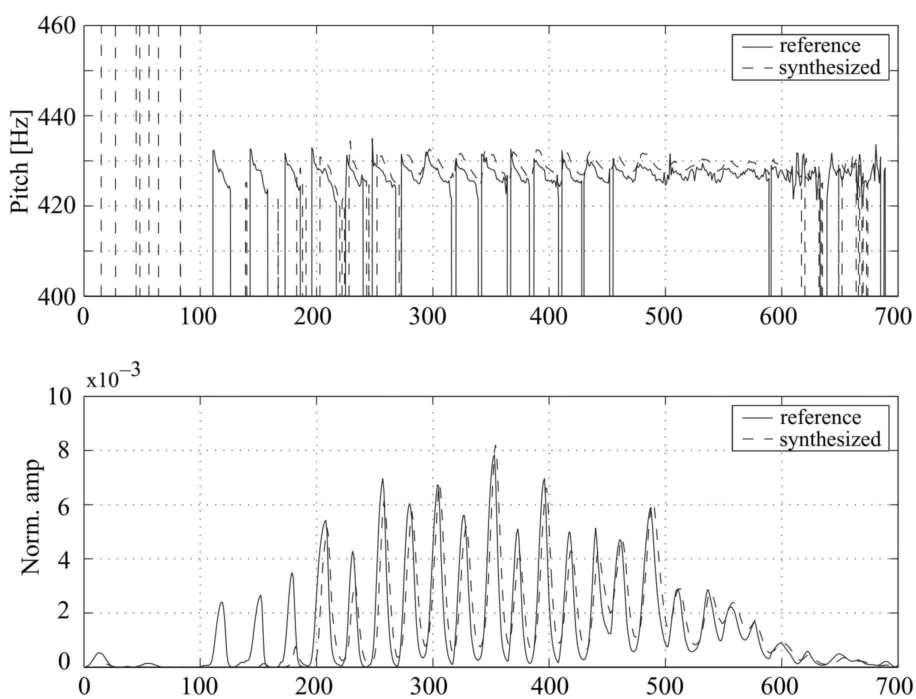

Fig. 7. Pitch contour and power envelope extracted from recorded and synthesized sounds (A4, tonguing).

employed in traditional music. The pitch contour and power envelope that were extracted from recorded and synthesized sounds are shown in Figure 7.

Although the power error seems to be relatively large, it shows a similar trend. This error is due to fast variation. Careful observation revealed that the power envelope of the synthesized tone lags behind that of the recorded tone. So, a time shift was permitted when the SDR was calculated. The result was 5.8 dB and the power correctness was found to be low. One reason is that the beginning part is missing in the case of synthesized tones. This is because the synthesis model has hysteresis characteristics. On the other hand, the pitch correctness was 98.5%.

We investigated the basic characteristics of the system using synthesized and recorded sho sounds, and determined the optimal parameter values for analysis experimentally. With fast amplitude-modulated sounds, there tended to be a delay compared with the reference.

3.3 Addition of various articulations using musical tones

This section describes the results when musical tones other than sho tones were used as a reference. Table 1 shows the type of articulations employed and the musical instrumental tones used in the experiment.

Table 1. Articulations and musical sounds used as reference signals.

Articulations	Musical sounds
Normal	Clarinet (Cl.)
Portamento	Hichiriki (Hc.)
Choking	Electric guitar (Eg.)
Vibrato (slow)	Flute (Fl.)
Vibrato (variable)	Soprano (Sp.)

3.3.1 No specific articulations

Figure 8 shows the pitch contour and power envelope of a clarinet tone (A4) as a reference, and also those of a synthesized tone. The results showed that sound with a similar pitch was synthesized and that the power envelope of the synthesized sound demonstrated the natural fluctuation of the original sound. However, we also observed an unnatural variation in the power envelope. Furthermore, some unnatural perceptual timbre variation was also noticed. This is because timbre and amplitude are affected by small variations in the physical parameters. At the beginning, it takes time before the oscillation builds up, and the power envelope is delayed compared with that of the reference. Objective scores for all the instrumental tones are summarized in Table 2 at the end of this section.

3.3.2 Frequency-dominant articulations

Portamento was analyzed as a second example. The hichiriki is an oboe-like double reed instrument that is normally played with relatively slow portamento. Figure 9 shows the pitch contour and power envelope of the original and the reference for the hichiriki. Generally, both curves are reproduced well, but a close inspection reveals a discrepancy. This is again because small variations in the physical parameters affect the threshold pressure of the oscillation, and hence the amplitude.

Choking is a technique mainly used in electric guitar playing. The result for a choking note is shown in Figure 10.

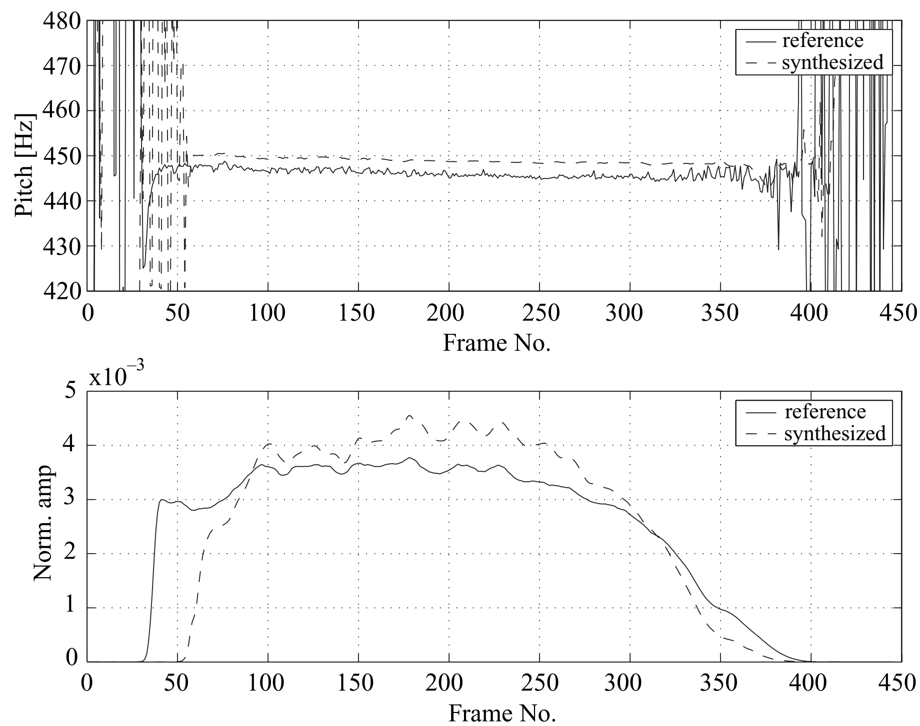

Fig. 8. Pitch contour and power envelope extracted from recorded and synthesized sounds (Cl., normal).

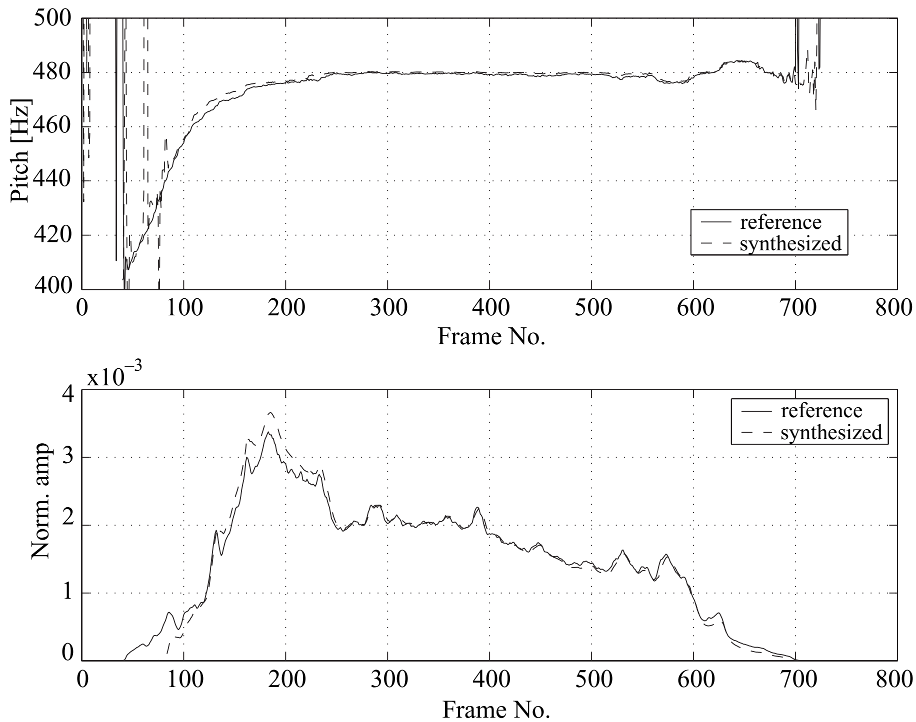

Fig. 9. Pitch contour and power envelope extracted from recorded and synthesized sounds (Hc., portamento).

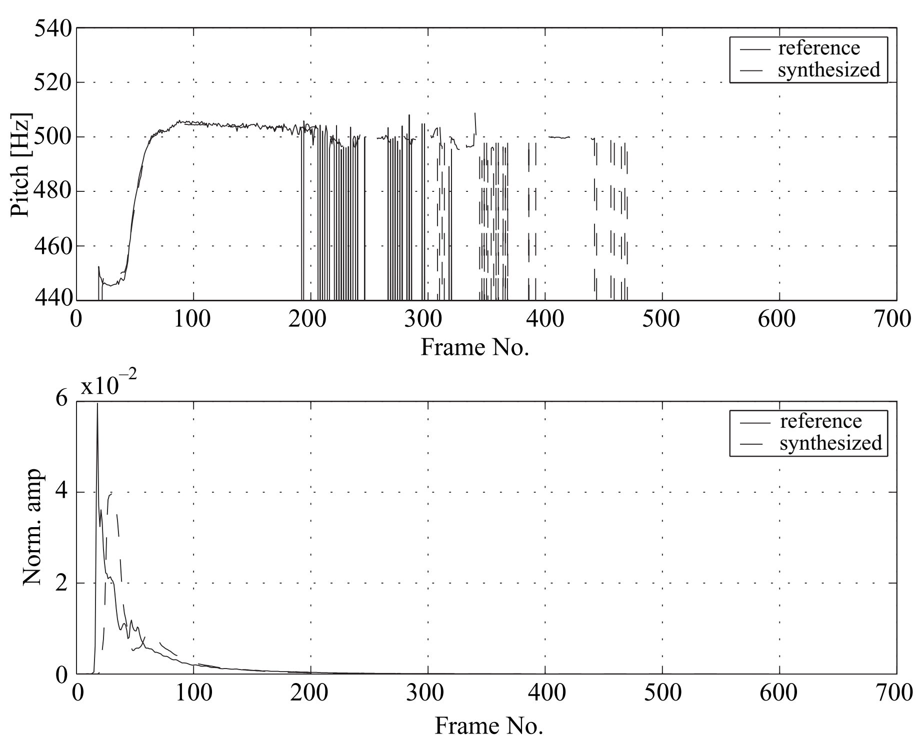

Fig. 10. Pitch contour and power envelope extracted from recorded and synthesized sounds (Eg., choking).

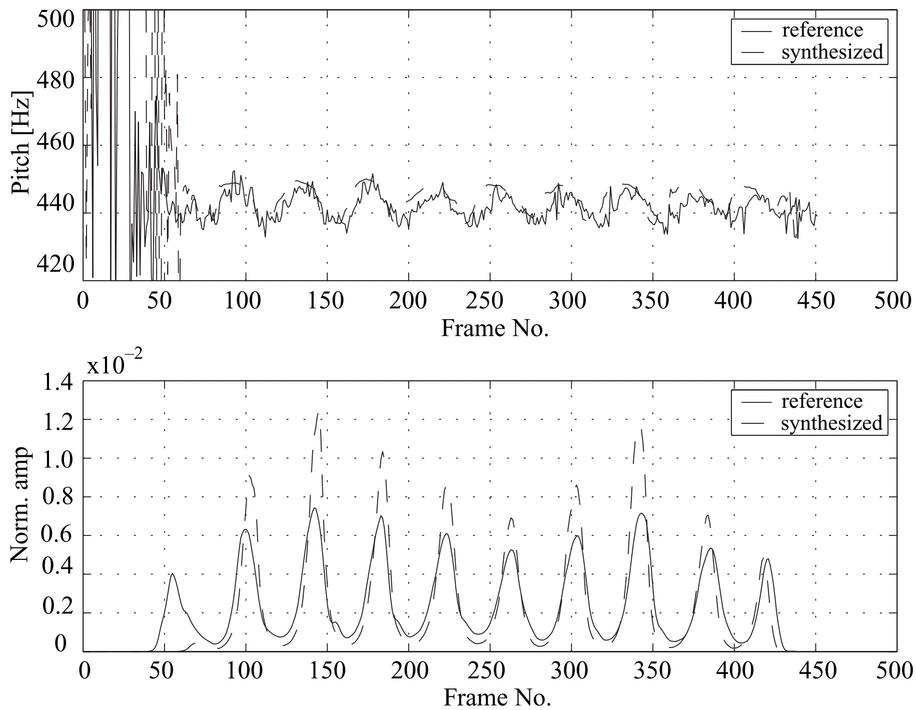

Fig. 11. Pitch contour and power envelope extracted from recorded and synthesized sounds (Fl., vibrato).

A considerable time shift in the power envelope was observed. This is because the power envelope of the original rises and decays so fast that the power envelope of the synthesized sound cannot catch up.

3.3.3 Vibrato

Figure 11 shows the pitch contour and power envelope of a flute note played with vibrato. Flute vibrato has both frequency and amplitude modulation. The pitch contours of the reference and synthesized sound correspond well. With the power envelope there is a larger discrepancy between the reference and synthesized sound, although a perceptually acceptable result was obtained.

Figure 12 shows results for a soprano voice with deep vibrato. The pitch contour of the synthesized sound agrees well with that of the reference even when there is a large pitch variation ranging between 420 and 500 Hz. In contrast, the power envelope exhibits degradation, although a perceptually acceptable result is obtained.

3.4 Discussion

Table 2 shows the objective score for each musical tone. The power correctness measured without and with a time shift were denoted as SDR1 and SDR2, respectively. The pitch correctness produced a fairly good result that exceeded 90% in most cases. In contrast, the power correctness result was not good, because of unexpected fluctuations and gross errors as already mentioned above.

Table 2. Pitch correctness and power correctness for various musical tones as a reference. SDR1 and SDR2 were calculated without and with time shift, respectively.

Reference	Pitch corr. [%]	Power corr. [dB]	
		SDR1	SDR2
Cl. normal	97.0	6.4	7.0
Hc. portamento	100.0	12.0	12.2
Eg. choking	97.3	1.7	5.2
Fl. vibrato	78.0	3.8	3.8
Sp. vibrato	93.6	5.8	5.8

An informal listening evaluation provided the following findings:

- Perceptually, sound quality degradation is less noticeable with a large pitch variation than with a small pitch variation.
- The objective criteria described in Section 3.1 do not necessarily correspond with a subjective judgment.
- There is a tendency for the rising part in the power envelope to delay, and sometimes stutters occur due to the physical properties of the model.

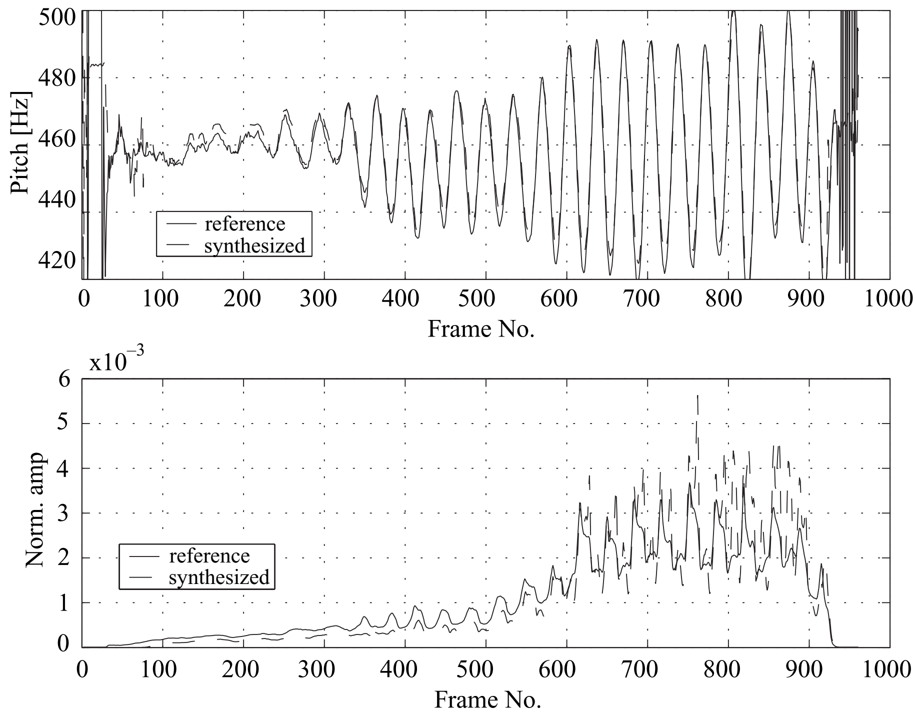

Fig. 12. Pitch contour and power envelope extracted from recorded and synthesized sounds (Sp., vibrato).

First, the effect of pitch error is discussed. There are small and large pitch errors and they affect performance differently; namely a small pitch error affects timbre and smoothness, and a large pitch error affects pitch perception. The former may be eliminated by employing post-process smoothing. We have obtained better perceptual results by smoothing, although the criteria did not improve.

An unexpected power envelope error occurred when the pitch contour of the input fluctuated. This error is inevitable in this current system, because power and pitch are treated separately in the analysis stage, whereas they are connected in the synthesis stage. One solution to this problem may be the use of another pitch table that represents the correspondence between the pitch and the (p, l, f) parameters.

As for the effect of the delay in the power envelope, there may be some difficulty when the proposed method is employed in a musical context and used to synchronize with other instruments. In Figure 7 the delay is about 10 ms, which corresponds to less than a sixty-fourth note in tempo = 60. Although this may not be negligible, it is small. In Figure 10, the delay is much larger for a guitar tone, and this corresponds to about 50 ms. This may become problematic. From the practical point of view, however, the delay can be adjusted after the synthesizing process because the current system is not designed for real time use.

In terms of actual use, a function that permits manual adjustment by the user is preferable. Therefore, we constructed a simple GUI to modify the control parameters.

4. Application to music composition

Other than the amplitude modulated and frequency modulated sounds mentioned above, more delicate and dynamic sounds can be obtained by carefully choosing parameters. To explore the possibilities provided by our system, a piece entitled “Morphing collage for piano and computer” was composed and premiered on 19 December 2002 at the Recital Hall, Tokyo Opera City. In this musical composition, in addition to its use as a simulator of the real instrument, the Sho-So-In system was used as a sound hybridization tool. The special trill and vibrato of the shakuhachi (Japanese bamboo flute) called “korokoro” and “yuri” were imitated, and sounds with sho timbre and shakuhachi articulations were produced by our system. Several segments of “korokoro” and “yuri” sounds were used as solo parts. Other sounds were used in chords as well. The features of the Sho-So-In were successfully introduced in the performance.

5. Conclusion

This paper described our synthesis method for sound hybridization based on a sho physical model. In accordance with this method, articulations were extracted from the given input signals, and sho-like sounds with these articulations were synthesized. This framework enables us to add more natural frequency and amplitude variations to model-based synthesized sounds. The system was further explored to pursue musically interesting timbres by modifying the parameters manually. Sounds created by this method were applied to a musical composition, and the method was shown to be effective. The sounds described in this paper can be heard by accessing the webpage http://www.kecl.ntt.co.jp/icl/signal/hikichi/jnmr/index.html.

Acknowledgments

The authors are grateful to Professor Ken'ichiro Ishii, a former Director of the NTT Communication Science Laboratories for his support, and Dr Keiji Hirata and Mr Ken-Ichi Sakakibara for fruitful discussions. Part of this work was supported by the Center Of Excellence (COE) formation program of the Ministry of Education, Culture, Sports, Science and Technology of Japan (No. 11CE2005).

References

Adachi, S., & Sato, M. (1995). Time-domain simulation of sound production in the brass instrument. *Journal of the Acoustical Society of America*, *97*, 3850–3861.

Burtner, M., & Serafin, S. (2002). The Exbow MetaSax: Compositional applications of bowed string physical models using instrument controller substitution. *Journal of New Music Research*, *31*, 131–140.

Caussé, R., Kergomard, J., & Lurton, X. (1984). Input impedance of brass musical instruments – Comparison between experiment and numerical models. *Journal of the Acoustical Society of America*, *75*, 241–254.

D'haes, W., & Rodet, X. (2003). A new estimation technique for determining the control parameters of a physical model of a trumpet. *Proceedings of the 6th International Conference on Digital Audio Effects*, 1–6.

Hélie, T., Vergez, C., Lévine, J., & Rodet, X. (1999). Inversion of a physical model of a trumpet. *Proceedings of the International Computer Music Conference 1999*, 149–152.

Hikichi, T., Osaka, N., & Itakura, F. (2002). A physical model of the sho and its application to articulation synthesis. *Proceedings of the International Computer Music Conference 2002*, 1–4.

Hikichi, T., Osaka, N., & Itakura, F. (2003). Time-domain simulation of sound production of the sho. *Journal of the Acoustical Society of America*, *113*, 1092–1101.

Mathews, M., Miller, J., & David, E. Jr. (1961). Pitch synchronous analysis of voiced sounds. *Journal of the Audio Engineering Society*, *33*, 179–186.

Moorer, J.A. (1979). The use of linear prediction of speech in computer music applications. *Journal of the Audio Engineering Society*, *27*, 134–140.

Noll, A.M. (1967). Cepstrum pitch determination. *Journal of the Acoustical Society of America*, *41*, 293–309.

Roads, C. (1996). Physical modeling and formant synthesis. *The computer music tutorial*. Cambridge, MA: MIT Press, 263–315.

Smith, J. (1996). Physical modeling synthesis update. *Computer Music Journal*, *20*, 44–56.

Tarnopolsky, A.Z., Fletcher, N.H., & Lai, J.C.S. (2000). Oscillating reed valves – An experimental study. *Journal of the Acoustical Society of America*, *108*, 400–406.

Tellman, E., Haken, L., & Holloway, B. (1994). Timbre morphing using the Lemur representation. *Proceedings of the International Computer Music Conference 1994*, 329–330.