# Sho-So-In: New Synthesis Method for Addition of Articulations Based on a Sho-type Physical Model

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

[1] NTT Communication Science Laboratories, NTT Corporation
[2] School of Engineering, Tokyo Denki University
[3] Graduate School of Engineering, Nagoya University
*email:* hikichi@cslab.kecl.ntt.co.jp

## Abstract

This paper proposes a synthesis framework that synthesizes sho-like sounds with the same articulations as the given input signal. This approach has three components: acoustic feature extraction, physical parameter estimation, and synthesis. At the acoustic feature extraction part, amplitude and fundamental frequency of the input signal were extracted, and the parameter estimation part converts them to control parameters of the physical model. Then, using these control parameters, sound waveform is calculated at the synthesis part. Based on this method, 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, portamento can be created. The system was successfully used in a music piece as a sound hybridization tool.

## 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 the real musical instrument. Our previous paper has shown that the proposed physical sho model has the same physical characteristics as 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.

Sho is categorized as Asian free-reed instruments, and this family of instruments has spread from east to south Asia. In Japan, sho is used to play chords or tone clusters in traditional gagaku music. Although many attempts have made to produce more dynamic and expressive sounds in contemporary music, there are limitations that arise from its 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 models can be modified without any loss of their timbral identity, and hence the models can be used to explore timbre. Here, we use this flexibility to implement articulations that we can find in other musical instruments, and attempt to extend the sho timbre space.

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). 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 Sho-type Physical Model

This section describes the configuration of Sho-So-In (stands for SHO SOunds INteresting), our sound synthesis system.

The main features are as follows. First, this system is expected to create sounds with naturalness, because model-based synthesis approach is applied. Second, being based on signal-based approach using real musical instrumental tones, precise control and specification are made possible.

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

- Acoustic feature extraction
- Physical parameter estimation
- Synthesis

The system configuration is shown in Fig. 1. Each part 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 are relevant to articulations. In this study, the fundamental frequencies and power per frame are used here as the most fundamental acoustic features. We will refer to the 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. In order to extract the pitch contour, cepstrum-based method is used.

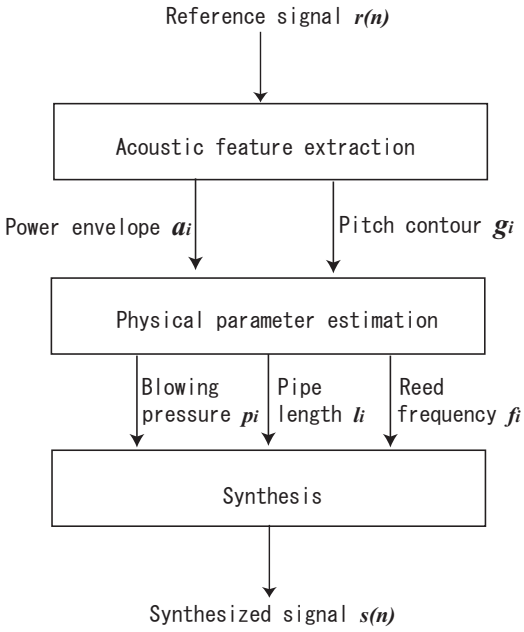


Figure 1. Configuration of the Sho-So-In system.

## 2.2 Physical Parameter Estimation

**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 sets for each pipe is 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. Purpose of controlling length $l$ here is not for on/off switching, but for more precise timbre and pitch control.

**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 $f$. Hence, the pitch table is created using synthesized sounds with different $l$ and $f$ pairs. Hikichi et al. (2003) showed that oscillation condition is satisfied when the resonance frequency of the pipe and the reed frequency have a certain relationship. Based on this knowledge, pairs of $l$ and $f$ are selected and synthesis is done.

1. Assume the fundamental frequency of a reference signal for *i*-th frame to be $g_i$, and search for the nearest frequency $\hat{g}_i$, and the second nearest frequency $\overline{g}_i$ from the pitch table.
2. Assume that the parameters corresponding to the fundamental frequencies $\hat{g}_i$ and $\overline{g}_i$ are $\hat{P}$ and $\overline{P}$, respectively. The parameter for the *i*-th frame is calculated by the interpolation of $\hat{P}$ and $\overline{P}$ at a ratio of distance from $g_i$.

**Determination of blowing pressure $p$.** Our preliminary investigation showed that the power envelope of the synthesized sound is roughly in proportion to 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 the low amplitude range is expanded. Hence, we assume the *n*-th root as the nonlinear mapping function from frame power to blowing pressure, and we determined 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 $p_i = P_{\max} \sqrt[n]{a_i}$, where $P_{\max}$ is the maximum pressure needed to adjust the absolute value.

According to the procedures described here, the control parameters for synthesis ( $p_i, l_i, f_i$ ) are specified for each frame time.

## 2.3 Synthesis

At the synthesis stage, control parameters are interpolated with time, and synthesis is done using our physical model of the sho. Only the basic equations are described. A more detailed derivation can be found in Hikichi et al. (2003).

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

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

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

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

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

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

By discretizing equations shown above, 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). Spherical radiation is assumed at the boundary condition. Using this transfer function, radiated pressure is calculated from the volume velocity obtained by the equations.

# 3 Experiments

## 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.

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

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

**Experimental Conditions.** At the acoustic parameters extraction part, acoustic features were extracted using a 20-ms window and a 5-ms shift, and pitch extraction and voiced/unvoiced discrimination based on cepstrum method were done. At the physical parameters estimation part, 30 pairs of ( $l$ , $f$ ) parameters with constant $p$ were used and synthesized sounds with fundamental frequencies ranging from 400 to 535 Hz were obtained.

The pitch contour and power envelope were extracted from the synthesized sounds in the same manner, and compared using the criteria mentioned above.

**Preliminary investigations of the system.** First, synthesized sho sounds were used as a reference signal.

Synthesized signals 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, pitch contour that changed linearly with time was provided manually, and the physical parameter was estimated and synthesis was performed. In this case, almost the same level of performance was obtained.

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

**Performance for recorded sho sounds.** Next, natural recorded musical sounds were applied to the system. To begin with, recorded sho sounds were used as a reference signal, and the power envelope and pitch contour were compared. The *n* value of the mapping function $p = P_{\max} \sqrt[n]{a}$ was used as a parameter.

The recorded sho sounds are naturally blown tones with no specific articulations. Their amplitude gradually increases, and decreases, and the duration is about 10 seconds.

The power envelope correctness shows a peak when *n* = 4 or 5. The pitch correctness was about 80 %. An informal listening test showed that the *n* value should not be made too large, because it would also make sounds in the silence 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.

The pitch contour and power envelope are plotted in Fig. 2. 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 recorded tone. The power envelope showed a nice correspondence.

It was found that voiced/unvoiced discrimination error occurred at the beginning and ending parts. This error is included in the pitch correctness criterion, and hence this criterion does not represent how well the pitch information is conveyed properly. To avoid this kind of error, only the frames that both the reference and the target were judged as voiced were taken into consideration. This modified criterion exceeds 99 %.

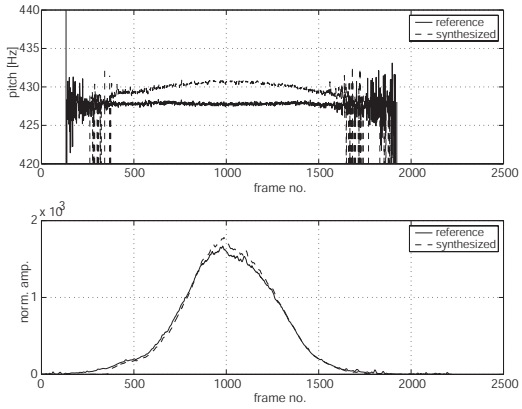


Figure 2. Picth contour and power envelope of the recorded and synthesized sho sounds.

## 3.3 Addition of Articulations using Musical Tones

This section describes the results when musical tones other than sho tones were used as a reference.

**Portamento.** Portamento tone was analyzed as a first example. The hichiriki is an oboe-like double reed

instrument that is normally played with relatively slow portamento. Figure 3 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 because small variations in the physical parameters affects threshold pressure of the oscillation, and hence the amplitude.

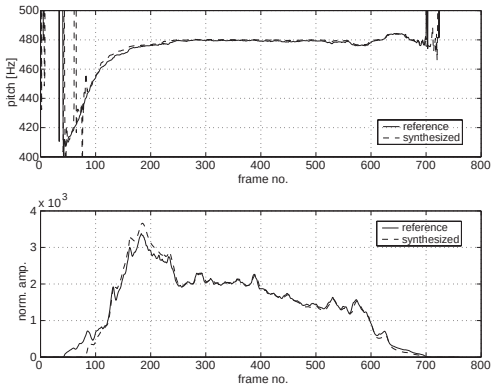


Figure 3. Result for a hichiriki portamento sound.

**Vibrato.** Figure 4 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.

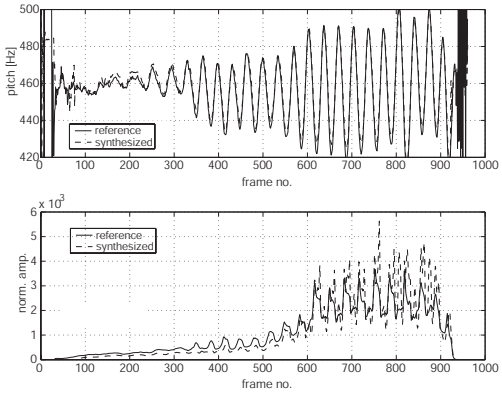


Figure 4. Result for a soprano vibrato sound.

### 3.4 Discussion

Generally, synthesized sounds have acceptably good quality in terms of articulation. However, some sort of error occurred. First, there is a tendency for the rising part in the power envelope to delay, and sometimes to disappear. An unexpected deviation also occurred when the pitch contour of the input fluctuated. This error is inevitable in the current system, because power and pitch are treated separately.

Informal listening evaluation demonstrates the following findings. 1) Perceptually, sound quality degradation is less noticeable with a large pitch variation than with a small pitch variation. 2) 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.

In terms of actual use, a function that permits manual adjustment by the user is preferable. This system also has manual mode, which control parameters can be adjusted manually through a simple GUI.

## 4 Use in a Music Piece

This system was used in music piece creation. The title of the piece is "Morphing collage" and was written for piano and computer, and was premiered on December 19, 2002 at the Recital Hall, Tokyo Opera City. Using Sho-So-In, sound hybridization was done with the shakuhachi (Japanese bamboo flute), besides the simulator of the real sho. A special trill of the shakuhachi called "korokoro" was imitated. This is a trill in which timbre is modulated and changes effectively while pitch remains fairly constant. Using the panel, three control parameters are controlled manually and sound hybridization was done. The features of Sho-So-In were successfully introduced in the performance.

## 5 Conclusion

This paper described our sound synthesis method for addition of articulations based on a sho physical model. Articulations were extracted from the given input signals, and sounds with these articulations were synthesized. This framework enabled us to add more natural expression to model-based synthesizers. The system was successfully used in a music piece as a sound hybridization tool.

## 6 Acknowledgments

Part of this work is supported by the Center Of Excellence (COE) formation program of the Ministry of Education, Culture, Sports, Science and Technology of Japan (No. 11CE2005).

## References

Hikichi, T., Osaka, N., and Itakura, F. 2003. “Time-domain simulation of sound production of the sho.” *J. Acoust. Soc. Am.* 113:1092-1101.

Hélie, T., Vergez, C., Lévine, J., and Rodet, X. 1999. “Inversion of a physical model of a trumpet.” *Proceedings of the International Computer Music Conference* pp. 149-152.

Hikichi, T., Osaka, N., and Itakura, F. 2002. “A physical model of the sho and its application to articulation synthesis.” *Proceedings of the International Computer Music Conference* pp. 1-4.

Caussé, R., Kergomard, J., and Lurton, X. 1984. “Input impedance of brass musical instruments – Comparison between experiment and numerical models.” *J. Acoust. Soc. Am.* 75:241-254.