CONSTRUCTION OF AN ELECTRONIC TIMBRE
DICTIONARY FOR ENVIRONMENTAL SOUNDS BY
TIMBRE SYMBOL

Yohei Kobayashi                    Naotoshi Osaka
Tokyo Denki University
School of Science and Technology for Future Life
2-2 Kanda nishiki-cho, Chiyodaku, Tokyo, 101-8457, Japan
yohei@srl.im.dendai.ac.jp, osaka@im.dendai.ac.jp

ABSTRACT
This paper introduces an electronic timbre dictionary, a
search or synthesis system for user desired timbre. Timbre is represented by timbre symbols newly defined
here. The method for designing timbre symbols is
discussed. The timbre dictionary is designed to open to
network users, so that these symbols can be defined and
edited by many of network users. Thus the number of
timbre symbols will increase. However, they are
expected to converge and become stable as entries in
Wikipedia have. The system is designed to ease
manipulation of timbre symbols and linked sound data
using a newly-defined graphical user interface (GUI)
interaction mode: snaked path listening. This paper
reports the total system configuration in detail. The
system is intended for use in automatic sound media
content creation and electronic music score-making and
performance.

1. INTRODUCTION
We have been studying a search and synthesis system
for environmental sounds for application to multimedia
content creation. We call the system an electronic
timbre dictionary. In order for users to directly
represent a sound image, a timbre symbol is newly
introduced, which represents timbre by symbol for a
short duration sound stream or sequences of such
symbols. We use timbre symbols as an input to the
system.

In order to represent a timbre, we have been
proposing a timbre symbols [6]. In reference [6], the
necessity for a timbre theory has been argued for, and
requirements for the theory have been discussed. We
stated that timbre symbols are indispensable, and the
design and structure of such symbols is discussed in
section 2.

In order to develop timbre symbols and make them
more sophisticated, we intend to release the dictionary
on the internet, so that any users can share it. The
timbre dictionary is not only referred to, but also
modified, edited and revised by many network users for
both timbre symbols and sound data, and resultantly
similarly to Wikipedia’s success, it is expected that it
will converge, become stable and reach high quality. In
section 3 and succeeding sections, its system design and
configuration are described.

2. DESIGN OF TIMBRE SYMBOLS
Let us classify known sound representation models or
synthesis systems from the viewpoint of the components
which they are composed of. We define those
components as timbre components, although not all of
them are perceived. Table 1 shows the classification of
timbre-component-based representations, which try to
resynthesize the original sound as closely as possible, in
terms of timbre components. The classification is
written in the order of granularity of the timbre
components, including additive synthesis. In No.1, a
sinusoidal model [7], sinusoids are used as timbre
component with the smallest granularity. SN ratio is the
measure of the fitness of the wave. In No.2, SMS [8],
noise is added to the timbre component. Noise is treated
as a stochastic process and justified since microscopically they are perceived in the same way, and reconstruction of a wave is not the goal of the system. These two components might not be appropriate to be called timbre components since their granularity is too small and cannot be perceived.

No. 3 is one of the target levels of our study, in which durations shorter than speech vowels are treated as a timbre component, and called micro-timbres. No. 5 is familiar to all, that is, onomatopoeia, whose duration is the same as that of speech phoneme. No. 4, musical instrumental sound, is a trial to represent in terms of phoneme, while in No.6, speech is expressed in terms of instrumental sound. In reference [10], a trial has been conducted to represent environmental sound such as birds singing in terms of MIDI sound sources. If timbre components of large granularity are adopted for representation of a sound, distortion increases even in an analysis/synthesis framework. These are useful for some artistic pieces expressing deformed sounds.

In this paper, timbre components are restricted to perceptive units of timbre as short as tens of milliseconds. This includes No.3 to 5. of the table. Timbre symbols are defined to represent all of these component: micro-timbre, phoneme, and onomatopoeia. We also define macro-timbres as a sequence of a few timbre symbols, which represent micro-timbre sequences or groups of those. Table 2 gives an example of timbre symbols for water sound. ‘p#’ represents water drop sounds in general. Figure 1 gives an example of micro timbres used as the lyrics. Once these symbols are defined for a sound, they are available to be used in writing a musical score. This will lead to new directions in the scoring of electroacoustic music.

This system is not suited for invariant textural sound such as various noises but is particularly useful for rapidly changing timbre such as the stirring of water and birdsong, as well as speech. Those symbols are initially provided with in a system. Users can add their own new definitions. For sounds which interest many users, symbols will converse among speakers of the same language. This optimistic thought is from the success of Wikipedia. Inversely, in the case of multiple sounds for one symbol, the system can browse all of them.

This framework provides us with infinite timbres. Timbres that we feel could use meaningful symbols for purposes other than use as sound sources for artistic music are: internal medicine stethoscopes, the struck sound of water melon to check its sweetness, and car engine sounds used to check for trouble.

3. IMPLEMENTATION OF ELECTRONIC TIMBRE DICTIONARY

Since timbre symbols are not familiar to users, the system needs to assist their timbre symbol manipulation. This section describes total system configuration in detail, including assistive technologies for timbre symbol interface such as several types of graphical user interface interaction modes: hierarchical type, vector type and snaked path listening.

3.1. Outline of the system

We implemented an electronic timbre dictionary. System functions and its relations are shown in figure 2. The timbre symbol database and wave database are shared by all the users via network. Network users can search and synthesize sounds from timbre symbols.

In a web browser of the client’s system, a user can access both the timbre symbol and wave database through text input and two methods of interaction with the GUI. The system has three modes:

![Table 1 Classification of timbre notation of sounds in terms of timbral components](image1)

<table>
<thead>
<tr>
<th>No</th>
<th>Timbre component</th>
<th>Granularity</th>
<th>Measure</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sinusoids</td>
<td>Small</td>
<td>SN ratio</td>
<td>Algorithm [7]</td>
</tr>
<tr>
<td>2</td>
<td>Sinusoids Noise</td>
<td>Small</td>
<td>SD</td>
<td>Algorithm [8]</td>
</tr>
<tr>
<td>3</td>
<td>Micro timbre</td>
<td>Medium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Phoneme</td>
<td>Large</td>
<td>SD</td>
<td>Algorithm [9]/Manual</td>
</tr>
<tr>
<td>5</td>
<td>Onomatopoeia</td>
<td>Large</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Instrumental sound</td>
<td>Large</td>
<td>SD/Perceptual distortion</td>
<td>Algorithm [9]/Manual</td>
</tr>
<tr>
<td>7</td>
<td>MIDI signal</td>
<td>Large</td>
<td>SD</td>
<td>Algorithm [10]</td>
</tr>
</tbody>
</table>

![Table 2 An example of timbre symbols for water sound](image2)

<table>
<thead>
<tr>
<th>Timbre component</th>
<th>Macro timbre</th>
<th>Onomatopoeia (Japanese)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drop</td>
<td>p#</td>
<td>picha, pisha, pita, pichi, pwan</td>
</tr>
<tr>
<td>Stream</td>
<td>ch#</td>
<td>choro, joro</td>
</tr>
<tr>
<td>Stir</td>
<td>sh#</td>
<td>Shower, shoer, jar, cop, gobo, pak, dok</td>
</tr>
</tbody>
</table>

![Figure 1 Example of micro timbre and its sequence](image3)
1. Sound data and timbre symbol registration mode
2. Timbre symbol editing mode
3. Sound looking up mode: search/synthesis

1 is an initial mode. Sound data are sent via network, and timbre symbols are defined together with other annotation and stored in the database. Figure 2 depicts a registration window. 2 is an editing mode. Supposing sound data have been stored and timbre symbols have already been defined. Symbols and their data can be edited by other users to improve the quality of the symbols and the database. 3 is the most important, and a primary purpose of the system. Other two functions are assistive ones for this looking up function. In this system, “looking up” a symbol in the dictionary does not only mean performing a sound search, but also synthesis if an appropriate sound does not exist.

In mode 3, there are two types of inputs: interactive input and batch input. Interactive input gives the server timbre symbols one by one using two methods provided in the GUI; this will be described in detail in the next section. In batch input, a timbre symbol script is input to the server. This directly links to search engine and accesses the timbre symbol database. It finally accesses the sound data which is linked to the timbre symbol.

3.2. Configuration of software system

In order for users to manipulate the system via web browser, a Flash is used as an interface on the basis of Adobe Flex Builder2 [11].
4.2. Snaked-path listening mode

In order for users to actively search for the timbres they want, we implemented a function which enables succeeding playback of sounds by tracing a path through the timbre icons shown in the mode introduced above. We call it snaked-path listening. Fig. 4 depicts the view for listening. It shows the traced path of a mouse cursor; sounds which are linked to these timber symbols on the path are played back in sequence. It also gathers touched icons and reassembles new timbres shown below the figure.

4.3. Sound synthesis functions

The system is also equipped with sound synthesis functions. These functions are convenient in the following cases:
1. a lack of sounds corresponding to the timbre symbol the user has input.
2. a comparison of a wave found in the wave database and a wave synthesized from given timbre symbols.
3. a sound data sequence defined by snaked-path listening.

In the case of 2, the user can choose between the sounds being compared. In the case of 3, the path defines a sequence of sounds. The synthesis functions can produce a wave file of those sounds played in sequence. At present, a limited number of synthesis methods are provided: STK (Synthesis Tool Kit) [12], physical-model-based synthesis engine of Sho [13][14], a Japanese traditional instrument and Karplus-Strong-model-based wavetable synthesis [15].

5. CONCLUSION

An electronic timbre symbol dictionary has been researched. It consists of both search and synthesis systems for user-desired timbres from timbre symbols for environmental sounds. In this study, some considerations for timbre symbols have been given, and a server-client system via network has been reported. We introduced two new types of GUI interaction modes: a hierarchical type and snaked-path listening. They make using the system more comfortable.

Our first version is in Japanese. However, we will endeavor to provide an English version, and hopefully we can release it to the public in the near future.

6. REFERENCES