AN ELECTRONIC TIMBRE DICTIONARY AND 3D TIMBRE DISPLAY

Naotoshi Osaka, Yoshinori Saito, Shinya Ishitsuka, Yasuhiro Yoshioka

Tokyo Denki University
School of Science and Technology for Future Life
osaka@im.dendai.ac.jp, {saito,ishitsuka,yoshioka}@srl.im.dendai.ac.jp

ABSTRACT

An efficient way to search for timbres is one of the most important functions of a sound manipulation tool for electroacoustic musicians and multimedia content creators. We are constructing an electronic timbre dictionary, which is a web-browser-based sound database search and synthesis system. The main characteristic of the system is that newly-defined timbre symbols are used to enable a reverse lookup of timbre. Timbre symbols are divided into three categories: macrotimbre, onomatopoeia, and microtimbre. These concepts are examined in comparison with previous research. Then the presently-implemented system is introduced, which incorporates the timbre symbols in terms of XML (timbre XML). A newly-developed 3D sound object display is also introduced.

1. INTRODUCTION

We have been constructing an electronic timbre dictionary, which provides users with instrumental sounds and environmental sounds in a web browser for the purpose of multimedia content creation [1]. By "providing" we mean making possible a search of a sound database and performing sound synthesis of the sound imagined by the user.

Among such systems, Freesound [2] is a representative one and used by many. Users can also register and evaluate sound with tags; as a result, the amount of information in the system is growing.

A sound search is done by either file name or other tags, and it uses a typical search style. However, if a creator uses such a system, it is desirable not to search for a sound source, but directly for the timbre a user wants. It is possible to search timbre by embedding timbre information in an annotation. However, this is indirect, and several trials have to be done in order to achieve the target timbre.

That is to say, Freesound is equipped with a forward search, while what many creators need is a reverse search. To enable the reverse timbre search, we introduce a symbol-based description of timbre into the system and have defined "timbre symbols."

We plan to release the electronic timbre dictionary as a browser-based server-client system in the near future. However, it is still under development, and a prototype system is being built, equipped with the fundamental functions needed to achieve the tasks stated above. Simultaneously, we are engaged in the study of timbre symbols. This paper describes the system concept, a timbre symbol design concept, and a GUI (Graphical User Interface) for sound object display.

2. DESIGN OF TIMBRE SYMBOLS

The system assumes that composers as users are manipulating various sounds in a timbre category for music or multimedia content creation. This is of interest to experts. For general purposes, the framework can also deal with general environmental sounds. The system takes both cases into consideration, but the new concept of timbre symbols is particularly defined for experts' use.

Such a system can also be applied to the problem of a "diagnostic analysis of artifacts in the sound of a car engine," "diagnostic analysis of health using a stethoscope by physicians", etc. Research details for those are out of the scope of this paper. However, one of the important common aspects for these applications is that the target sounds are of interest to and are identified only by the experts. The design of timbre symbols is intended to be appropriate in such cases.

2.1. Environmental Sound Recognition Using Onomatopoeia

Onomatopoeia are ready-made timbre symbols, and it is natural to incorporate them in the system. Japanese, the authors' mother tongue, is rich in onomatopoeias: a Japanese onomatopoeia dictionary has several thousand entries, and introducing those seems to be appropriate.

In facing the design of timbre symbols, the problem of multiplicity arises: how a sound is heard as an onomatopoeia differs from listener to listener. Ishihara et al. give a solution to this problem [3][4]. They first experimentally showed for a single-syllable sound that the "process of transforming environmental sounds into syllable structure expressions is listener-independent," and that expression of onomatopoeia is listener-dependent [3]. In order to solve the ambiguity problem, they define a Basic Phonetic Group (BPG) which includes a pair of phonemes used frequently for an identical environmental sound by different listeners.

They also experimentally showed that this strategy works better rather than making a broader articulatory group such as fricative consonants [4]. For example, /t/ and /k/ are in the same group, while /p/ is in a different one. A high recognition rate was reported using BPG. In this research, general environmental sounds are dealt with conscious of the requirements for robot hearing.

2.2. Timbre Symbols and Its Conversion

Timbre component	Macro timbre	Onomatopoeia(Japanese)
Drop	p#	picha, pisha, pita, pichi, pwan
Stream	ch#	choro, joro
Stir	sh#	shower, shoer, jar, cop, gobo, puk, dok

Table 1. An example of timbre symbols for water sound

Timbre symbols are defined according to granularity: macrotimbre, onomatopoeia and micro-timbre. This is a classification based on auditory perception, and differs from categorical classification based on sound generation described in the next chapter. For convenience of explanation, we start with describing onomatopoeia.

1. Onomatopoeia: This completely adopts the phonetic description system of a user's mother tongue. Onomatopoeia and mimetic words are not independent. An onomatopoeia might also convey the atmosphere from the sound and may not completely reflect the acoustic reality. The onomatopoeia used here completely reflects how sounds are heard by the listener. Initially BPG has been adopted for symbols.

2. Microtimbre: This is setup for the purpose of verifying similar sounds for ordinary subjects but different sounds for experts. In order to design symbols which will be familiar to a large number of users, the symbols are based on IPA (International Phonetic Alphabet) initially, and user-defined options might be possible in the future.

3. Macrotimbre: Contrary to microtimbre, this represents some meaningful unit, such as water drop sounds as a whole, and the bird song of one specific species. One code for one group is given. There are two types of macrotimbre: one is a group which represents sound generation or specific events regardless of timbre. A group of breathing sounds caused by bronchial pneumonia might be defined as a single symbol. The other is completely dependent on how it is heard. Water drop sounds and the pizzicato of a string instrument can be heard as highly similar and be represented using a single macrotimbre.

We assume the representation of sound which has a syllable structure. However, for stationary noise such as machine noise or electroacoustic noise with a certain texture which does not have such a structure, we also apply the timbre symbol system stated above. Table 1 gives an example of timbre symbols for water sounds. 'p#' represents water drop sounds in general.

Freesound can be said to be a Wikipedia of sound. Similarly, the electronic timbre dictionary is a Wikipedia of timbre symbols as well as sound. In the cases of Wikipedia and Freesound, the content quality is much higher than first imagined and converges. We believe the change of timbre symbols will follow a similar path.

We also consider another aspect: this system provides users with the potential of giving birth to a new language. Twin talk, which is created by twins when very young and is not comprehended by their mother, and the existence of dialects and various foreign languages in sign languages proves that a small number of people can produce a stable language.

2.3. Application of Timbre Symbol

Music systems and theory, as well as linguistic systems, consist of discrete structures that can be represented as symbols, such as interval, scale, chord and tonality. Common music notation is a form of music representation in terms of these discrete symbols. This symbolic description enables the recording, transmission and preservation of music. On the other hand, because there is no descriptive system for new timbres, it has been difficult to convey music to others in an intermediate status so that it can be performed. It is only possible to record, preserve and convey a full piece music in a perfect digital archive. As a result, people other than the composer can relate to that music only as listeners and never as "performers" of the computer part. In this research, a trial has been made to represent all the timbre in a piece of music as symbols (scripts), particularly for non-instrumental sounds, such as environmental sounds and electro-acoustic sounds.

3. IMPLEMENTAION OF ELECTRONIC TIMBRE DICTIONARY

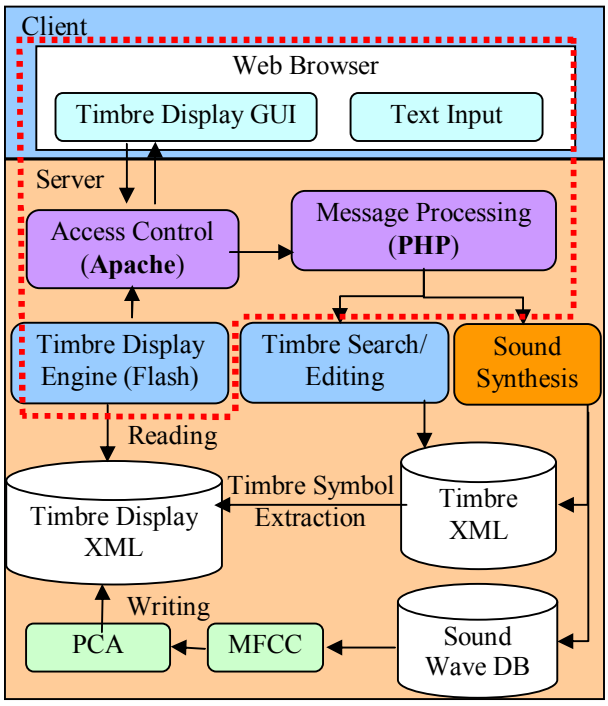
Figure 1 depicts the system functions and their relationships. From a web browser a user can access the timbre symbol database in the system server through GUI and text input. Presently implemented functions include sound registration, timbre symbol list display, editing timbre symbols and timbre symbol search.

In registration mode, a user can register timbre symbol, keyword, comments and other information together with environmental sounds.

Figure 2 shows an example of a sound and a sound timbre list. Sound play back from the list is also possible.

Searching for not only timbre symbols but also keywords is possible.

In order for users to manipulate the system via web browser, a Flash-based GUI was made with Adobe Flex Builder2 [5].



* DB: Database • • • : Interface

Figure 1. Block diagram of system functions

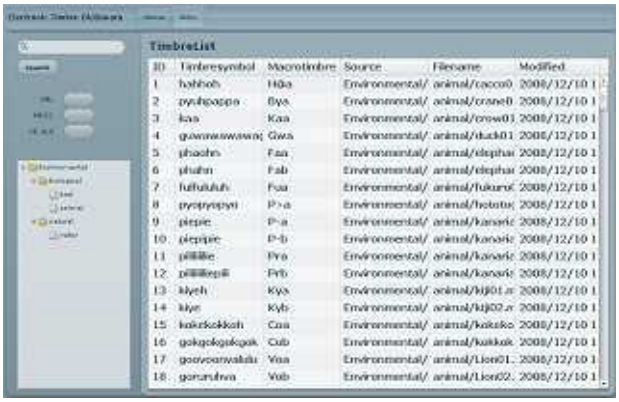


Figure 2. Sound list with timbre symbols and annotation

. The web server runs on Windows. Apache is used to control access from web browsers on client system. In order to decrease the load to clients, a script language, PHP, is used server-side to process requested messages.

The synthesis control component is written in C#, and the synthesis engine is equipped with a dynamic-link library (DLL).

3.1. Hierarchical Structure of Timbre

In general, timbre symbol definition is done after sound registration. This also happens when one records sounds and registers (uploads) them, and another defines timbre symbols for them later. Then it is convenient to distinguish the hierarchical structure of timbre symbols from that of sound in the physical sound database. In order for users to manage sound data independently from timbre symbols, hierarchical structures based on the generative aspects of the sounds have been defined, rather than perceptive aspects.

In the top level, two categories are set up: instrumental sounds and environmental sounds. The second layer of environmental sounds includes natural sounds, life sounds, animals, etc. At present, our sound data is limited; it includes 39 animal sounds, 318 percussion sounds, and 198 water-stirring and drop sounds.

In order to use timbre symbols as annotation to an electronic timbre dictionary, an XML format has been adopted to describe timbre symbols, and timbre XML is newly defined. Timbre XML is a text based timbre description corresponding to a wave file.

4. TIMBRE DISPLAY INTERFACE

In relation to timbre search, references [6] and [7] are well known as audio browsing tools. Trials have been done to design a sophisticated aural and visual display that maps sound files of interest to assist in sound analysis, synthesis, and retrieval problems.

In order to synthesize compound sounds from a large sound database efficiently using sound browsing, real-time corpus-based concatenative synthesis (CBCS) [8] is known. Sound is visualized, and sound search, playback and modification become possible for the purpose of music creation. CBCS enables 1) two-dimensional display of sounds, 2) representation of each sound as a dot, and 3) real-time synthesis and playback of the mouse-covered area. However, real-time viewpoint change is not possible.

Besides automatic timbre search, manual exploration is also important for sound manipulation. The next section discusses the requirements of the visual display to explore various sounds efficiently.

4.1. Necessary Functions for Sound Display and Manipulation

Taking perceptive characteristics of timbre into consideration, the next characteristics are added to the visual symbols for sound data (sound object).

1. A sound object exists and is displayed independently for a sound (sound stream).
2. Sound objects are placed in timbre-metric space, and a good correspondence between perception and displacement is obtained.
3. Real-time manipulation through the display is possible.
4. Sound playback is the highest-priority function.

One spherical particle corresponds to a sound stream and is displayed in a 3D space. A spherical object display is

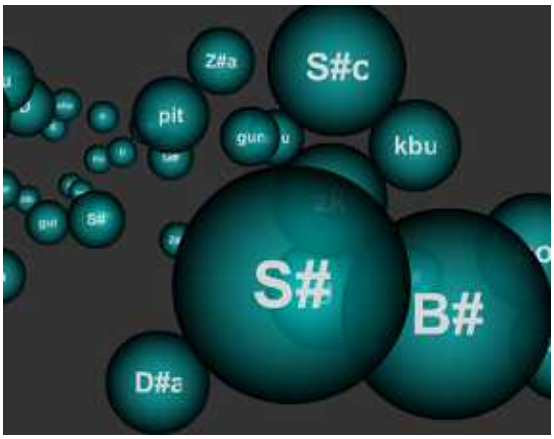


Figure 3. An example of sound display interface

very familiar to many users since it is a common macroscopic-to-microscopic model, used in space scenery, solid state physics models such as molecules and atoms, and biological models. Exploring such a field might give users a game-like sensation. The loudness, pitch and texture of a sound can correspond to the size, colour, shape and surface shape of a particle. For items 2 to 4, the 3D display of a sound clarifies the sound manipulation by matching physical scope to timbre perception.

In order to satisfy item 3, affine transform such as space movement, zooming and widening of scope, and rotation of viewpoint are possible by using a mouse as if the user were moving among heavenly bodies by means of a rocket.

4.2. Sound Display Method and Conversion to 3D Data

In order to display sound objects, 24-dimensional MFCC (Mel-Frequency Cepstral Coefficients) are derived and the dimensions are reduced to three using Principal Component Analysis (PCA).

To enable sound objects to be expressed in 3D space, sound objects are represented in terms of XML again, and timbre display XML is newly defined. This is produced by extracting relevant data from the timbre XML, and adding MFCC data to it. After reading the timbre display XML, a sound object is displayed as a particle together with the timbre symbol on its surface. Figure 3 is an example of the display of multiple sound objects.

4.3. Discussion

At present the limit of the number of simultaneous sound objects is 50. If the number exceeds 50, real-time manipulation becomes difficult. For a wider view, the display method of CBCS, using a small dot for each sound, might be a better solution. Our informal test results showed that if the 3D GUI is used, the time needed to search for a target sound from eight instruments is reduced to 74% of that when performing the search by list alone (Figure 2).

5. CONCLUSION

The recent progress of the development of an electronic timbre dictionary, which is a Wikipedia-like database of sounds particularly aimed at experts, has been reported. Timbre symbols are defined and discussed in comparison with previous research. Using timbre symbols for the reverse lookup of sound becomes possible. It is equipped with the fundamental functions: registration of sounds and timbre symbols, modification, and timbre search. A 3D sound object display has also been implemented as a new GUI to the system, which is believed to enable efficient exploration of sounds corresponding to the users' interest.

Since the framework of a timbre-symbol-based reverse lookup itself is quite new, several steps must be undertaken before releasing the system onto the network. Our next step is to distribute the system to other researchers or composers to get feedback on improving the system. We also plan to do experiments to investigate the characteristics of timbre symbols.

6. ACKNOWLEDGEMENT

We are sincerely grateful to Renick Bell for our enthusiastic discussion. This work was supported by KAKENHI 20520134.

7. REFERENCES

[1] Kobayashi ,Y. and Osaka, N. "Construction of an electronic timbre dictionary for environmental sounds by timbre symbol," Proc. of the ICMC, Belfast, 2008.

[2] Music Technology Group, "The freesound Project," http://freesound.iua.upf.edu/

[3] Ishihara, K. et al. "Automatic transformation of environ-mental sounds into sound-imitation words based on Japa-nese syllable structure," EUROSPEECH-2003, pp. 3185-3188, 2003.

[4] Ishihara, K. et al., "Disambiguation in determining phone-mes of sound-imitation words for environ-mental sound recognition," INTERSPEECH-2004, pp. 1485-1488, 2004.

[5] Adobe, "Flex Builder2," http://www.adobe.Com/jp/products/flex/

[6] Tzanetakis, G., and Cook, P., "3D Graphics tools for sound collections," Proc. Of the COST G-6 Conf. On DAFX-00, Verona, Italy, 12, 2000.

[7] Brazil, E., Fernstroem, M., Tzanetakis, G., and Cook, P., "Enhancing sonic browsing using audio information retrieval," ICAD02, Kyoto, Japan, 2002.

[8] Schwarz, D., et al. "What next? Continuation in Real-time corpus-based concatenative synthesis," Proc. ICMC 2008, Belfast, 2008.