

Voice Drummer: A Music Notation Interface of Drum Sounds Using Voice Percussion Input

Tomoyasu Nakano[†]

Masataka Goto[‡]

Jun Ogata[‡]

Yuzuru Hiraga[†]

[†]Graduate School of Library, Information and Media Studies, University of Tsukuba
1-2 Kasuga, Tsukuba, Ibaraki 305-0821, Japan
{nakano, hiraga}@slis.tsukuba.ac.jp

[‡]National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
{m.goto, jun.ogata}@aist.go.jp

ABSTRACT

Voice Drummer is a percussion instrument notation interface using spoken percussion patterns as input. Its internal framework is based on the authors' previous study on voice percussion recognition. *Voice Drummer* can be used in various applications such as composition, arrangement, and score editing. The system has the capability of adapting to individual users during the practice sessions, enabling performance improvement and also adding entertainment value. A trial usage of the system shows that it would be a useful tool for users with little knowledge or experience in percussion instruments.

ACM Classification H5.2 [Information interfaces and presentation]: User Interfaces. - Voice I/O; H5.5 [Information interfaces and presentation]: Sound and Music Computing.

General Terms Performance, Human Factors

Keywords: Voice percussion recognition, drum pattern retrieval, music-notation interface

1 INTRODUCTION

Notating music is difficult for those without substantial musical knowledge or skill. Such potential users can be supported by use of transcription systems that convert music played or hummed by the user into appropriate representations. Another aspect of user support is to provide visual/audio interfaces that are intuitively easy to understand and manipulate. Previous work on music transcription has focused on the extraction of pitch and duration, which is insufficient for percussion instruments where recognition of timbre and rhythm patterns are more dominant.

Voice Drummer is a percussion instrument notation interface, built as an application of the authors' previous work on voice percussion recognition [6]. The user sings out a drum pattern (*voice percussion*), which is analyzed and matched with entries in a drum pattern database, based on onset timing patterns and intended drum types. The retrieved pattern is the result of input recognition. Currently, *Voice Drummer* recognizes expressions of bass drums and snare drums.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

UIST'05, October 23-27, 2005, Seattle, Washington, USA.
Copyright ACM 1-59593-023-X/05/0010 ...\$5.00.

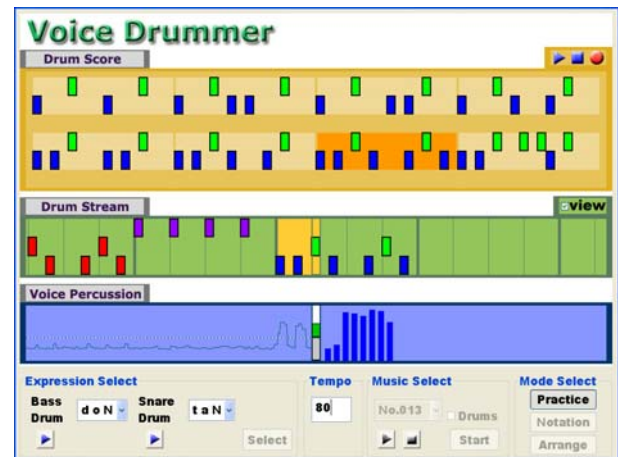


Figure 1: An example *Voice Drummer* screen (practice/adaptation mode).

Voice percussion in our context are verbal expressions such as *don-don* or *ta-ta*, which are transcribed into phonemic representations (or *onomatopoeia*). Such verbal form is more familiar and easier for the average user to express, as compared to trying a faithful acoustic imitation of actual drum sounds. In that sense, *Voice Drummer* is intended to be a useful tool for users with little knowledge of or experience with percussion instruments. Direct tapping used in rhythm games such as “*TAIKO: Drum Master*” are alternative methods of input, although the task will become more difficult when handling multiple drum types.

Voice Drummer also provides graphical representations and audio feedback of the recognition results, operating interactively with the user in real time. This interface enables easy understanding and manipulation of the results, and also adds a game-like, entertainment quality. The snapshot of an example *Voice Drummer* session is shown in Figure 1.

2 OVERVIEW OF VOICE DRUMMER

Voice Drummer operates in three modes — *practice/adaptation mode*, *notation mode*, and *arrangement mode* (described below). The modes are selected by clicking the corresponding buttons in the lower right portion of the *Voice Drummer* screen.

The upper part of the screen consists of the following three windows.

- **“Drum Score” window**
This window shows the graphical score of drum patterns (e.g. recognized results for voice percussion) presented in two rows with four measures each. The blue rectangular markers in the lower line represent onset of bass drums, and the green markers in the upper line the onset of snare drums.
- **“Drum Stream” window**
Drum patterns (lower two lines) and cue sounds (upper line: hi-hat sound) are dynamically shown as a real-time flow, moving from right to left. The vertical bar at the center corresponds to “present time”, so the left half shows past events, and the right half, future (anticipated) events.
- **“Voice Percussion” window**
Input power level is shown in the colored indicator at the center of this window, which flows out to the left as a line graph. The vertical bars to the right show the evaluated adaptation achievement in the practice/adaptation mode.

The exact meaning and purpose of the windows vary according to the operation mode.

2.1 Selection of Onomatopoeia

Bass drums and snare drums are distinguished by their onomatopoeic expressions and their timing in the rhythm patterns. The onomatopoeic expressions are looked up in a pronunciation dictionary, which can be limited to a smaller vocabulary using the “Expression Select” menu in the lower left portion of the screen. This in general will improve the processing speed and recognition rate.

2.2 Practice/Adaptation Mode

This mode runs practice sessions for the user to practice voice percussion input. The user first selects one of the 8 practice patterns in the Drum Score window (1 measure each). The selected pattern is run in the Drum Stream window, with a lead cue of 4 beats in a hi-hat sound. After hearing a sample play, the user performs voice percussion on the timing when the drum markers pass the “present time” bar.

Besides practice itself, the system adjusts its internal parameters to the vocal characteristics of the user in order to improve the recognition rate (*adaptation*). The likelihood of the recognition (adaptation achievement) is displayed on the right half of the Voice Percussion window. The saturation of its growth indicates when to finish the practice.

2.3 Notation Mode

In this mode, the drum patterns sung by the user are analyzed, recorded and displayed by the system. The input is taken one measure at a time. The recognized results are successively added to the Drum Score window, and also played back. The entire result can be replayed or stored in SMF format.

The system can also play stored drum patterns in random order in the Drum Stream window, which will help the user in idea generation or learning various drum patterns. The stored drum patterns are extracted from the RWC Music Database (Popular Music) [1].

2.4 Arrangement Mode

In this mode, the user’s task is to provide the drum part by voice percussion to a performance of melodic instruments. The input piece is performed by units of 8 measures, and the

user sings out the intended drum pattern. The recognized results are displayed in the Drum Score window, and can be played back with the melodic parts. The results can be stored in SMF format.

The input piece is selected by the “Music Select” menu in the lower right of the screen. If the original data has drum tracks, the drum sounds can be played or erased, together with the flow display on the Drum Stream window.

3 INTERNAL MECHANISM OF VOICE DRUMMER

(Verbal) voice percussion is an indirect, metaphoric expression of drum patterns, with prominent individual differences in both the vocal quality of and also the use of phonemic (onomatopoeic) expressions by the user. Thus, the existing methods used for acoustic drum sound recognition [3] or BeatBoxing (faithful acoustic imitation) recognition [2, 4] are not applicable in our setting.

In our study [6], we took the approach of using onomatopoeia as internal representation of drum sounds, and defined the recognition of voice percussion as the retrieval of intended drum patterns. Specifically, the system uses a pronunciation dictionary of onomatopoeic expressions of bass drums and snare drums using the data obtained from expression experiments and following the framework of speaker independent acoustic model provided by CSRC [5]. It has been experimentally confirmed that the recognition rate of this acoustic model can be improved through adaptation with only a limited number of sample voice percussion data (18 patterns).

4 CONCLUSION

This demonstration presented Voice Drummer, a music notation interface using voice percussion recognition. More detailed and live information of the system can be seen on the video presentation accompanying this presentation.

Trial usage of the system suggest that it would be a useful tool for users with little knowledge and/or experience in percussion instruments. Remaining issues such as the inclusion of instruments other than bass drum and snare drum are topics of future work.

REFERENCES

1. Goto, M., Hashiguchi, H., Nishimura, T., and Oka, R. RWC Music Database: Popular, Classical, and Jazz Music Databases. In *Proc. of ISMIR 2002*, pp.287–288, 2002.
2. Hazan, A. Towards Automatic Transcription of Expressive Oral Percussive Performances. In *Proc. of IUI 2005*, pp.296–298, 2005.
3. Herrera, P., Yeterian, A., and Gouyon, F. Automatic Classification of Drum Sounds: A Comparison of Feature Selection Methods and Classification Techniques. In *Proc. of ICMAI 2002*, pp.69–80, 2002.
4. Kapur, A., Benning, M., and Tzanetakis, G. Query-by-Beat-Boxing: Music Retrieval for the DJ. In *Proc. of ISMIR 2004*, pp.170–177, 2004.
5. Lee, A., Kawahara T., Takeda K., Mimura M., Yamada A., Ito A., Itou K., and Shikano, K. Continuous Speech Recognition Consortium – an open repository for CSR tools and models –. In *Proc. of LREC 2002*, pp.1438–1441, 2002.
6. Nakano, T., Ogata, J., Goto, M., and Hiraga, Y. A Drum Pattern Retrieval Method by Voice Percussion. In *Proc. of ISMIR 2004*, pp. 550–553, 2004.