DEVELOPMENT OF MIDI ENCODER “Auto-F” FOR CREATING MIDI CONTROLLABLE GENERAL AUDIO CONTENTS

Toshio Modegi

Research & Development Center, Dai Nippon Printing Co., Ltd.

250-1, Wakashiba, Kashiwa-shi, Chiba, 277-0871 Japan e-mail: Modegi-T@mail.dnp.co.jp

Abstract: The MIDI interface is originally designed for electronic musical instruments but we consider this music-note based coding concept can be extended for general acoustic signal description. At first we proposed applying the MIDI technology to coding of bio-medical auscultation sound signals such as heart sounds. Then we have tried to extend our encoding target and improve the coding precision based on Generalized Harmonic Analysis in order to apply it to vocal sounds. Currently, we are trying to separate each tone included in popular songs and encode both vocal sounds and several background instrumental sounds into separate MIDI channels. Using a GM-standard MIDI tone generator, we can playback this multi-channel MIDI encoded data including vocal sounds. In this paper, we present an abstract algorithm of our developed MIDI software encoder tool, being used for producing interactive general audio contents controlled by MIDI.

Key words: MIDI coding, audio contents, automatic notation, acoustic signal processing

1. Introduction

MIDI (Musical Instrument Digital Interface) is originally designed for musical instrument, and we are considering MIDI as an ideal coding method because of its coding efficiency and high-quality sound reproduction capability. The first application of MIDI technology was synthesising musical acoustic signals, especially generating rhythmic pattern signals substituting such as drum instrument set. Therefore, the first trial of perception of music signals was also for rhythmic patterns [1]. As features of

---

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: [10.1007/978-0-387-35660-0_65](https://doi.org/10.1007/978-0-387-35660-0_65)

R. Nakatsu et al. (eds.), *Entertainment Computing*

© IFIP International Federation for Information Processing 2003

the MIDI coding are similar to those of text formats, if it is applied to audio databases, we can retrieve audio contents by audio keywords or music-note strings [2].

We have been interested in multimedia medical databases, especially audio databases for heart sounds and lung sounds, and we have proposed the MIDI encoding method especially for heart sounds, and this algorithm features its real-time processing capability [3]. Besides our implementation of this proposed method for a heart-sound coding, we tried applying a MIDI coding to other types of sound materials including bird singing voices [4]. And we found out this converted MIDI data could be used for new types of interactive audio contents producing non-existing natural sounds.

In order to process various types of acoustic signals, we categorise general acoustic signals as two groups whether the spectrum components are distributed intermittently or continuously. Most natural acoustic signals including human voices and biological signals belong to the latter continuous group, whereas musical acoustic sounds except percussion sounds belong to the former intermittent group. Then we defined two kinds of MIDI coding approaches depending on processing complexity: a real-time coding method and a high-precision coding method [4].

As a result of our implementation of both types of coding, using the high-precision coding method, we found out it was possible to playback speeches and singings by MIDI tone generators, using our proposed non-linear extended GHA (Generalized Harmonic Analysis) frequency analysis method [5]. Then we have focused on the decoder sound module, and tried to produce more natural sounds as the original PCM sounds [6]. We also improved the frequency analysis precision by variable frame-length analysis, evaluated coding precision [7].

These days we have been developing sound source separation techniques, especially separation of vocal parts from mixed down songs. Considering our proposed method as a very low bit-rate audio codec, we have evaluated and compared it with the other encoding methods. We have reported the encoding quality of 8-kbps data by our MIDI encoding method was superior to that of 16-kbps MPEG-1 layer 3 encoded data [8].

Furthermore, using these improved coding techniques we are trying to apply this MIDI coding to symbolic expression of acoustic signals for retrieving music archives by note-based keywords. As a structured symbolic description format, we choose XML (eXtensible Mark-up Language)[9] because this format is widely used for medical application [10]. In this paper, we are going to overview our improved MIDI encoding algorithm, which has been implemented into Windows software and is now distributed for free. We expect this software tool to be used for producing interactive general audio contents controlled by MIDI.

2. Encoding Method

2.1 Background and Concept of MIDI Coding

These days more and more MIDI applications are created such as *karaoke*, cell phone pager, music game contents, pet robots, automatic piano and player's guide data for music keyboards. In order to provide existing or newly released music record contents to these applications, sound format conversion processes are necessary. Today these conversion business needs are expanding especially in Japan. However these processes are dependent totally on human ears called “ear-copy” and skilled manual operations, and they need trained musical talent. As a solution several MIDI direct input devices imitating musical instruments are being devised by several musical instrument makers, but these devices restrict musical performances. Therefore an automatic conversion tool including automatic notation processing is being expected.

The general application of MIDI technique called DTM (Desk Top Music) is that digitising music score written by a composer into the MIDI format and producing music instrument parts on a desk top without musician, musical instrument nor recording studio. And this technique is widely applied to today's commercial music production. However, for non-musical acoustic materials such as vocal parts, which are difficult to be expressed with MIDI music notes, you need singers and a recording studio facility.

In this sense, we proposed the other way, which converts the existing audio waveform materials to MIDI music notes. In this method you can control any kind of audio materials by MIDI functions interactively and reproduce even vocal sounds by MIDI tone generators or electronic musical instruments. And editing converted MIDI codes, we can also reproduce the similar music scores written by composers.

Our proposing MIDI coding is a kind of analytic-synthetic coding approaches which separate a given audio signal to amounts of sinusoidal waveforms and describe them with the frequency and intensity parameters of their separated harmonic sine waveforms. Whereas our method separates them to several predefined harmonic complex waveforms, which available MIDI tone generators can generate, and describes them with the frequency (namely MIDI defined *note-number*) and the amplitude (namely MIDI defined *velocity*) parameters of their separated predefined harmonic complex waveforms. In general the number of required harmonic complex waveforms for describing will be not so many as those of the analysed sinusoidal waveforms because each harmonic complex waveform is made of several sinusoidal waveforms. Therefore, the coded bit-rate of this MIDI method

will become smaller about 1/10 than that of the previous general analytic-synthetic coding approaches.

2.2 Two Types of Approaches for MIDI Coding

MIDI data are a collection of pairs of *Note-On* and *Note-Off* command strings called events where each pair denotes a piece of music note, and each event is composed of a relative time-stamp (*delta-time* in MIDI standard terms), a frequency (*note-number*) and an amplitude (*velocity*) parameter [3]. In this section we describe how these MIDI parameters can be numerically obtained. As we described, we proposed two approaches depending on the types of source acoustic signals whether musical acoustic or the other signals.

Using a frequency analysis technique such as GHA (Generalized Harmonic Analysis) method [5], we can separate some frame $g(t)$ (frame-length= T) extracted from the given acoustic signal. By the variable frame-length analysis technique [7], we can obtain a set of N separated sinusoidal functions as follows:

$$g(t) \cong \sum_{n=1, N} \{ A_n \sin(2\pi f_n t) + B_n \cos(2\pi f_n t) \} . \quad (1)$$

Where the both coefficients of A_n and B_n are defined by the following equations.

$$A_n = \sum_{t=0, T_n-1} 2/T_n \{ g(t) \sin(2\pi f_n t) \} . \quad (2)$$

$$B_n = \sum_{t=0, T_n-1} 2/T_n \{ g(t) \cos(2\pi f_n t) \} . \quad (3)$$

In these equations T_n is the maximum value of $T_n = k/f_n < T$ (k : appropriate positive integer value), and f_n is given by the equation $f_n = 440 \cdot 2^{(n-69)/12}$ which generates frequency values on the MIDI *note-number* logarithm scale. Defining harmonic complex functions as $u_i(t)$, we can rewrite the equation (1) with smaller summation elements $P \ll N$ as follows:

$$g(t) \cong \sum_{i=1, P} \alpha_i u_i(t) . \quad (4)$$

Then we define $p(i)$ as a representative frequency identification number of $u_i(t)$. In the first case like shown in Figure 4-(A), $u_i(t)$ can be expressed with a summation of a fundamental frequency $f_{p(i)}$ and its harmonic components $j f_{p(i)}$, as follows (j : integer value 1, 2, 3...):

$$u_i(t) = \sum_{j=1,J} \{A_i(j)\sin(2\pi j f_{p(i)} t) + B_i(j)\cos(2\pi j f_{p(i)} t)\} . \quad (5)$$

In the other case like shown in Figure 4-(B), $u_i(t)$ can be expressed with a summation of a formant local peak frequency $f_{p(i)}$ and continuous frequency components distributed in its neighbour $\beta_j f_{p(i)}$, giving the integer value of δ around 1, as follows ($f_{p(i)-\delta} \leq \beta_j f_{p(i)} \leq f_{p(i)+\delta}$) :

$$u_i(t) = \sum_{j=1,J} \{A_i(j)\sin(2\pi \beta_j f_{p(i)} t) + B_i(j)\cos(2\pi \beta_j f_{p(i)} t)\} . \quad (6)$$

If we choose the harmonic complex function $u_i(t)$ from the wave tables defined in our using MIDI tone generator, we can reproduce $g(t)$ with P number of notes giving *note-number* N_i and *velocity* value V_i . These values are generated from $f_{p(i)}$ and α_i parameters respectively as follows:

$$N_i = 40 \log_{10} (f_{p(i)}/440) + 69 . \quad (7)$$

$$V_i = 128 C \alpha_i^{1/2} \quad (C: \text{constant}, 0 \leq V_i \leq 127) . \quad (8)$$

The time of *Note-On* of this MIDI note event is the start position of extracted frame on the source acoustic signal, and the duration time (or *Note-Off delta-time*) is given by the analysed frame-shifted interval τ .

Figure 4. Two types of approaches for MIDI coding

2.3 Algorithm Design of MIDI Coding

Figure 5 shows the whole process of our designed MIDI encoding process. The first part is the Frequency Analysis, which separates the part of a source signal to N number of spectrum components and needs the most calculation load. The next Harmonic Grouping of Notes is integrating selected multiple spectrum components or notes, but in our implementation this process is just selecting out representative notes by their volume value.

Figure 5. Abstract flowchart of MIDI encoding process

The first two processes will be repeated by shifting the extracted position until the end of the sound file. In case the source signal is stereo, two sets of notes are analysed for each extracted position, and integrated to a set of notes with pan-pot parameters added. The fourth Temporal Grouping of Notes is connecting the temporally adjacent notes, which have similar frequency and volume parameters, and producing a longer duration note.

The last two processes are converting each integrated note to the MIDI event data format. Before that we should regulate the number of output notes or output bit-rate in order that the standard GM or other types of MIDI tone generator can play-back encoded data.

The right side flowcharts in Figure 5 show the several utilisation processes after MIDI data are created. The top three functions of MIDI Data Editor, Common Music Notation Tool, and MIDI Sequencer can be provided

in commercially available off-the-shelf DTM (Desk Top Music) composition tools such as “Yamaha XG-Works” or “Steinberg Cubasis” what we use. Not specifically stated in this paper, we are also considering structuring and symbolising the encoded MIDI data into XML (eXtensible Markup Language) document format for network audio content distribution.

Figure 6. A snapshot of MIDI encoder software tool

3. Conclusions

In this paper, we have described an abstract MIDI encoding algorithm based on constructing harmonic complex functions from the sinusoidal waveforms analysed by GHA. Extending this algorithm, not specifically described in this paper, we could also separate vocal parts from singing song materials and encode both vocal and instrumental parts into multiple channel MIDI data streams. And we could generate complete musical sounds including vocal sounds with a single GM-standard MIDI tone generator. Figure 6 shows an encoding example of piano and vocal separation; the source audio material was “Irish folksong: Danny Boy” and its length was

20 seconds. The output bit-rate was 10 kbps and its calculation time was about a minute using a Pentium III 600 MHz Windows98 PC.

As for future works, we are considering higher accurate sound source separation techniques, removal techniques of harmonic overtone components for automatic music notation, support techniques of pitch-bend functions for improving decoded sounds, higher performance of structuring and symbolising techniques for generating XML data, and an algorithm redesign for real-time processing.

This research has been promoted by the Digital Content Association of Japan as a 2000-year government project: "Development of Multimedia Content Creating Tools," being also financially supported by the Information-technology Promotion Agency Japan and the Ministry of Economy, Trade and Industry Japan. The developed software MIDI encoder tool (currently Japanese MS-Windows edition only !) is distributed for free at the following Web site. (URL: <http://www.dcaj.or.jp>)

References

- [1] M. Goto and Y. Muraoka, "A Beat Tracking System for Acoustic Signals of Music," *Proceedings of ACM international conference on Multimedia*, pp.365-372, 1994.
- [2] R. J. McNab, L. A. Smith, I. H. Witten, C. L. Henderson and S. J. Cunningham, "Towards the Digital Music Library: Tune Retrieval from Acoustic input," *Proceedings of the 1st ACM International Conference on Digital libraries*, pp.11-18, 1996.
- [3] T. Modegi and S. Iisaku, "Application of MIDI Technique for Medical Audio Signal Coding," *Proceedings of IEEE 19-th International Conference of the Engineering in Medicine & Biology Society*, Chicago, pp. 1417-1420, Oct. 1997.
- [4] T. Modegi and S. Iisaku, "Proposals of MIDI Coding and its Application for Audio Authoring," *Proceedings of International Conference on IEEE Multimedia Computing and Systems*, Austin, USA, pp.305-314, Jun. 1998.
- [5] T. Modegi, "Multi-track MIDI Encoding Algorithm Based on GHA for Synthesizing Vocal Sounds," *Journal of Acoustic Society of Japan (E)*, Vol.20, No.4, pp.319-324, 1999.
- [6] T. Modegi, "High-precision MIDI Encoding Method Including Decoder Control for Synthesizing Vocal Sounds," *Proceedings of the seventh ACM international conference on Multimedia*, Part 2, Orlando, USA, pp.45-48, Nov. 1999.
- [7] T. Modegi, "MIDI Encoding Method Based on Variable Frame-length Analysis and its Evaluation of Coding Precision," *Proceedings of IEEE International Conference on Multimedia & Expo*, New York, pp.1043-1046, Aug. 2000.
- [8] T. Modegi, "Very Low Bit-rate Audio Coding Technique Using MIDI Representation," *Proceedings of ACM 11-th NOSSDAV Workshop*, New York, pp.167-176, Jun. 2001.
- [9] T. Modegi, "Structured Description Method for General Acoustic Signals Using XML Format," *Proceedings of IEEE International Conference on Multimedia & Expo*, Tokyo, Japan, pp.932-935, Aug.2001.
- [10] T. Modegi, "XML Transcription Method for Biomedical Acoustic Signals," *Proceedings of 10th World Congress on Health and Medical Informatics Medinfo2001*, London, UK, pp.366-370, Sep.2001.