

# Rencon 2004: Turing Test for Musical Expression

Rumi Hiraga<sup>\*</sup>  
Bunkyo Univ.  
1100 Namegaya,  
Chigasaki  
253-8550, Japan

Roberto Bresin<sup>†</sup>  
Royal Inst. of Technology  
Drottning Kristinas, v. 31  
SE-10044 Stockholm,  
Sweden

Keiji Hirata<sup>‡</sup>  
NTT Comm. Sci. Lab.  
2-4 Seika-cho, Hikaridai  
Souraku-gun, Kyoto  
619-0237, Japan

Haruhiro Katayose<sup>§</sup>  
Kwansei Gakuin Univ.  
Sanda, Hyogo  
669-1337, Japan

## ABSTRACT

Rencon is an annual international event that started in 2002. It has roles of (1) pursuing evaluation methods for systems whose output includes subjective issues, and (2) providing a forum for researches of several fields related to musical expression. In the past, Rencon was held as a workshop associated with a musical contest that provided a forum for presenting and discussing the latest research in automatic performance rendering. This year we introduce new evaluation methods of performance expression to Rencon: a Turing Test and a Gnrut Test, which is a reverse Turing Test, for performance expression. We have opened a section of the contests to any instruments and genre of music, including synthesized human voices.

## Keywords

Rencon, Turing Test, Musical Expression, Performance Rendering

## 1. INTRODUCTION

Rencon (Contest for Performance Rendering System) is an annual international event that started in 2002. We have pursued evaluation methods for performance rendering systems through Rencon. Its goal is to foster research on computational models of and methods for the generation of expressive musical performances. In the past, Rencon was held as a workshop associated with a musical contest that provided a forum for presenting and discussing the latest research in automatic performance rendering and, more generally, computer-based music performance and its expression.

Regardless of the instrument, a performer interprets a musical piece and generates musical expression. This process of understanding and interpreting a piece, deriving a performance plan, and expressing it musically has been the focus of performance rendering systems. Many instruments with new user interfaces presented at NIME will also be played according to the same process. Moreover, musical performance is a promising field for investigating basic principles of human emotion, intelligence, creativity and individuality. This is the same posture of Minsky's awareness of issues on music as human activity [11]. To answer the broad,

<sup>\*</sup>rhiraga@shonan.bunkyo.ac.jp

<sup>†</sup>roberto@speech.kth.se

<sup>‡</sup>hirata@brl.ntt.co.jp

<sup>§</sup>katayose@ksc.kwansei.ac.jp

complicated, and common question of what is musical expression, Rencon proposes common evaluation ground for system-rendered performances.

## 2. PERFORMANCE RENDERING SYSTEMS

Before introducing the details of Rencon, we describe and classify performance rendering with computer systems. Studies on performance rendering have been published since the 1980s. In those days, automated performances caught people's attention simply because they were new. When people noticed the monotonousness of system rendered performances, the systems introduced musical structures that resided in a musical score and utilized implicit knowledge and a performance plan to improve musical expression.

Performance rendering consists of three stages: (1) preprocessing where musical analysis or learning performance occurs, (2) performance rendering, and (3) post-processing of manually modifying expressions of rendered performances [4]. According to the degree of human intervention in each of the above stages, we classify performance rendering with computers into three types: (1) manual rendering, (2) assisted, and (3) autonomous. The manual rendering has only the third stage of performance rendering.

Both assisted and autonomous types are software systems that automatically generate expressive performances from a musical sheet and information that is unique to each system. Post-processing is prohibited. While assisted type provides users with better usability and richer musical information than sequence software, the assisted type allows the arbitrary involvement by users at the preprocessing stage and follows a specific performance for learning. The ultimate style of autonomous type is based on learning models which can automatically analyze music and select the most suitable rendering for a given score. Since we have no complete and satisfying automation of music analysis such as GTTM, even autonomous type uses manually derived information.

## 3. RENCON HISTORY

### 3.1 The Dawn

In Japan in the mid 1990s, more and more people began to notice computer music research because of the expectations of the multimedia environment and the new research area of *Kansei*. After the International Computer Music Conference in Japan in 1993, a community of computer music research was officially established. In those days, each performance

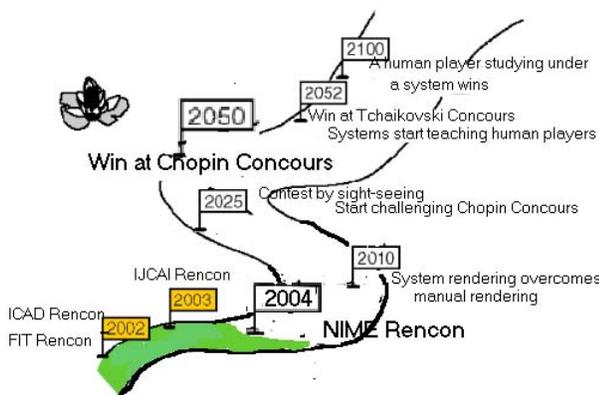


Figure 1: Road map of Rencon

rendering system was evaluated subjectively by different sets of criteria for individual music pieces, or showed the reproduction of musical expression in a target music using the information learned with a model performance.

A music contest for system rendered performances was proposed in 1996 to encourage research [10]. In 1997, at the IJCAI workshop “Issues on AI and Music – Evaluation and Assessment” [2] the evaluation of computer music systems was the common issue. At a domestic symposium in 2000, six performance rendering systems gave demonstrations [7]. Many people noticed that listening and comparing performances is helpful for understanding different systems.

We started a working group to make the listening comparison into practice [6], and posed a road map of Rencon activity outside the computer music research community [5] (Figure 1). The slogan “Winning at the Chopin Concours in 2050” represents Rencon as a landmark project whose value lies in developing technologies and understanding music as a human activity towards the winning process.

### 3.2 Rencon in the Past

Rencon was born in 2002. So far, there have been three Rencon workshops. We introduced new trials for listening comparison in each workshop, set musical sections for different purposes of listening comparison, and used a common sound source (Table 1).

#### ICAD Rencon

Rencon started as a satellite workshop of the International Conference on Auditory Display (ICAD) 2002 in July in Kyoto (ICAD Rencon). It was an entire day of workshop consisting of eight technical presentations, a general discussion on the common basis for performance rendering contest, and listening comparisons accompanied by the public voting which based simply on listeners’ subjective preference for a performance. As the first Rencon, any genre of piano performances were allowed. All six performances, one manually rendered and five by systems, were played on a Yamaha Disklavier grand piano. First prize went to a manually rendered piece for which about eighty percent of audience showed their preference. The second prize went to *Director Musices* by Bresin [1].

#### FIT Rencon

In September 2002 in Tokyo we held the second Rencon as a special event of the Forum on Information Technology (FIT Rencon), which is the biggest academic IT forum in Japan. The free, half day workshop was open to the public and sought to enlighten people about computer music, performance rendering, and Rencon. Musical pieces were restricted to compositions by Mozart or Chopin; ten performances (one was manually rendered) were played. Each performance chose its own hardware sound generator. Performances were evaluated by the audiences and a music specialist. According to both the audience and the musical expert, the winner was *MUSE* (assisted type) by Taguti [15]. It showed that “system rendering overcomes manual rendering by 2010” target (Figure 1) arrived earlier than we expected. Moreover, many people admitted that system-rendered performances were actually rather good [8].

#### IJCAI Rencon

The third Rencon was an official workshop of the International Joint Conference on Artificial Intelligence in August 2003 in Acapulco entitled “Methods for Automatic Music Performance and their Applications in a Public Rendering Contest” (IJCAI Rencon). During the one day workshop, there were eleven presentations and two contests consisting of a compulsory section restricted to pieces by Chopin and an open section. A unique sound source of GIGA Piano was used for both the compulsory and open sections. Six pieces were entered in the compulsory section and four in the open section. The winners of the compulsory section were *MUSE* and *Raphael’s* [13] (assisted type). *MUSE* also won the open section.

### 3.3 Analysis of Past Rencon Results

We offer a few observations and comments from past Rencon workshops.

#### Commonality in Music Appreciation

After FIT Rencon, a set of performances was repeatedly used for listening comparison. The audience reactions to the listening comparisons have revealed a common tendency in appreciating music. More than three hundred people of different ages, musical backgrounds, and ethnicity participated in the listening comparisons. These comparisons always showed the same grouping of performances by their orders<sup>1</sup> regardless of the constituent of audiences. We should also notice the judgements by the audience and an expert at FIT Rencon were the same [3].

#### Human Intervention

The results of past Rencon workshops suggest that currently the human support and intervention in rendering performances generates the more appreciated music. This relates to the commonality issues of the contest which we describe next.

#### Toward Comparing what a System Rendered

Simply put, evaluating performance rendering systems seeks to clarify and compare the deviation for expression that each

<sup>1</sup>For example, the top three performances are the same for any listening comparisons although the order changes.

	ICAD Rencon	FIT Rencon	IJCAI Rencon	NIME Rencon
Musical section	none	none	Compulsory and Open	Compulsory (Turing Test)/ Open/Gnirut Test
Compulsory music	none	Chopin's or Mozart's	Chopin's	Chopin's
Common sound source	YAMAHA Disklavier	none	GIGA piano	GIGA piano
Judgement	audience voting	audience voting and evaluation by a specialist	audience voting	audience voting

Table 1: Comparison of past Rencon workshops

system appended during the rendering process. Therefore it is desirable and necessary that rendered pieces are performed and listened to under the same conditions. Identical conditions increase the ease with which listeners compare and judge music, resulting in more reliable evaluations. For that purpose, Rencon has established compulsory music and a common sound source.

Rencon will also provide an XML based music database for case and compulsory music as a Rencon entry kit [9][12]. The entry kit includes the score, performance, and the deviation data<sup>2</sup>, allowing researchers to concentrate on their systems. The rendered piece also reflects more directly the technical points in handling deviation for expression.

Under these conditions and without arbitrary human intervention, what we listen to during Rencon music comparison may represent the technical aspects of each system. Therefore, we also believe the autonomy should be encouraged and human intervention in rendering a performance should be discouraged. As will be described in 4.2, we ask each system the degree of autonomy at Rencon 2004 in NIME.

#### 4. RENCON 2004 IN NIME

Many listeners have acknowledged that past performances at Rencon were “good”<sup>3</sup>. Thus it is not premature to introduce Turing Test type of listening comparison in Rencon.

##### 4.1 Turing Test for Musical Performances

The Turing Test was proposed by A. Turing to judge whether a program has intelligence after a conversation with a human being [16]. If a person believes the other party is also a human being, then the program is regarded to have intelligence. Although most AI researchers disapproved of judging intelligence in such a manner, a restricted Turing Test, called the Loebner Prize Competition, started in 1990 [14]. At the competition, human judges conversed with agents who consisted of contestants (natural language systems) and confederates (human beings), and rated each agent for their human-likeness. Even though many researchers continue to doubt the validity of the competition, it has continued.

The Turing Test for musical performances determines by listening whether system-rendered performance is distinguishable from human performance. At past Rencons, we found a commonality in music appreciation among people of different musical backgrounds. The Turing Test may reveal

<sup>2</sup>A musical performance is never identical to the notes on a score. During performance such factor as onset time and offset time may modify the information on the score.

<sup>3</sup>A good performance does not necessarily give people a deep artistic and aesthetic impression.

Please complete the following questionnaire.

Does your system analyze a musical piece automatically?	My system is able to analyze a musical piece <input type="radio"/> fully automatically <input type="radio"/> semi-automatically <input type="radio"/> in a sense yes, in a sense no <input type="radio"/> slightly automatically <input checked="" type="radio"/> not at all
Do you need to manually modify your performance after your system has generated it?	My system can complete a musical piece <input type="radio"/> without any human intervention <input type="radio"/> with little human intervention <input type="radio"/> in a sense with, in another sense no human intervention <input type="radio"/> with significant human intervention <input checked="" type="radio"/> with much human intervention
Does your system use parameters for performance at the high-abstract level, such as the time-span-tree, or the low-abstract level, such as a MIDI parameter itself?	My system uses parameters for performance, and <input type="radio"/> all of them are high-abstract <input type="radio"/> most of them are high-abstract <input type="radio"/> in a sense high, in other sense low-abstract <input type="radio"/> some of them are low-abstract <input checked="" type="radio"/> all of them are low-abstract
What is the degree of generality of your system?	It is <input type="radio"/> very high <input type="radio"/> high <input type="radio"/> medium <input type="radio"/> low <input checked="" type="radio"/> very low
How long did it take you to render the submission piece?	About <input type="text"/> Min., <input type="text"/> Sec.

Figure 2: Questionnaire for participating systems

different aspects of music as a human activity. In judging whether a performance is rendered by a human player, a person must understand musical expression systematically<sup>4</sup>. During a Turing Test, audiences will be required to listen to music more actively.

##### 4.2 The Autonomusness

As described above, we would like to judge by listening what a system rendered (added information as expression). Thus human intervention in any stage of performance rendering obscures the scope of what a system accomplished. At submitting musical pieces, we asked each system for its autonomy, generality, etc. (Figure 2). Though the data are not reflected in the listening comparisons, these questions will form an evaluation method by making a relationship between technological superiority and generated music in the future.

##### 4.3 Three Sections of Musical Contest

The Rencon musical contest is made of three sections: compulsory, open, and a Gnirut test. The compulsory section aims to pursue common ground for evaluating performances

<sup>4</sup>Whether a listener can explicitly describe the understanding is different from understanding musical expression.

rendered by computer systems. The open section and the Gnirut test were planned to increase the appeal of Rencon for listeners who are not experts in music performance.

### Compulsory Section

The compulsory section will specify rigid commonality in playing musical data. Music is restricted to compositions by Chopin. Each performance rendering system generates musical data in SMF that will be performed at the contest using GIGA Piano as the common sound source. The compulsory section will be held in Turing Test style; listeners will judge whether a musical expression is generated by a system or a human. Musical performances will be anonymously performed in random order, then the listeners will rate performances by humanlikeness without being informed of the number of each performance. They will also indicate their preferences. A prize called the Rencon award will be given to the performance generated by a system evaluated as being the most human-like. In the case of a tie, audience preference will be considered.

### Open Section

We hope that the open section will increase interest in systems for expressive performance rendering. No musical genres, instruments, or sound sources are excluded. Chamber music, a piano piece, and three songs by synthesized human voices will be presented in this section. The audience will listen to performances played anonymously and evaluate each one by the same evaluation method as in past Rencon listening sessions.

### Gnirut Test

This is another type of Turing Test where machine-ness is judged by audiences. Music submission is restricted to a short piece composed by J. S. Bach. GIGA Piano will be used as a common sound source. The submissions are performed and judged in a similar way to the compulsory section.

## 5. CONCLUDING REMARKS

In the past, Rencon provided good opportunities for participants to listen to and compare musical expression generated by different performance rendering systems. This year again listeners will actively participate in Rencon listening comparisons. So far, more than a few people agree that some performance rendering systems generate music that rivals human performances. Thus, introducing the Turing Test and Gnirut Test to Rencon is timely.

Piano pieces by Chopin are not easy for performance rendering systems and few researchers who work on performance rendering have attempted with his pieces. Besides considering the Chopin Concours, we included Chopin in the compulsory section because of its difficulty. Once a performance rendering system is able to render them impressively, especially with autonomy, then many mysteries of music expression will be solved.

The latest information on Rencon events, contest results, and sound data is on the Rencon WEB page:

<http://shouchan.ei.tuat.ac.jp/~rencon/index.shtml>

## 6. REFERENCES

- [1] R. Bresin and A. Friberg. Emotional coloring of computer-controlled music performances. *Computer Music Journal*, 24(4):44–63, 2000.
- [2] K. Hirata. (ed.). *IJCAI 97 Workshop Issues on AI and Music Evaluation and assessment*. IJCAI, 1997.
- [3] R. Hiraga, R. Bresin, K. Hirata, and H. Katayose. After the first year of rencon. In *Proc. of ICMC*, pages 135–138. ICMA, 2003.
- [4] R. Hiraga, M. Hashida, K. Hirata, H. Katayose, and K. Noike. Rencon: toward a new evaluation method for performance rendering systems. In *Proc. of ICMC*, pages 357–360. ICMA, 2002.
- [5] R. Hiraga, K. Hirata, and H. Katayose. Rencon: the ambitious pianist! In *Transaction of IPSJ*, pages 136–141. IPSJ, 2002.
- [6] R. Hiraga, Y. Horiuchi, T. Murao, and Y. Takeuchi. Rencon: a workshop for planning a piano contest by performance rendering systems—workshop report and panel discussion. In *2001-MUS-41*, pages 37–42. IPSJ, 2001.
- [7] R. Hiraga, H. Katayose, H. Koike, T. Suzuki, K. Noike, and T. Hoshishiba. Performance rendering 2000: Demonstration and panel discussion. In *2000-MUS-35*, pages 67–70. IPSJ, 2000.
- [8] R. Hiraga, C. Oshima, and K. Nishimoto. What’s cooking in rencon? In *2003-MUS-50*, pages 19–24. IPSJ, 2003.
- [9] K. Hirata, K. Noike, and H. Katayose. Proposal for a performance data format. In *IJCAI2003 workshop program, APP-5*, pages 65–69. IJCAI, 2003.
- [10] H. Katayose, M. Goto, Y. Horiuchi, T. Matsushima, T. Murao, S. Shimura, T. Rai, and K. Hirata. Report on the panel discussion: computer music research as computer science. In *96-MUS-15*, pages 91–98. IPSJ, 1996.
- [11] M. Minsky. Music, mind, and meaning. *Computer Music Journal*, 5(3), 1981.
- [12] K. Noike, K. Hirata, and H. Katayose. A report of the rencon-kit, the 1st release. In *2003-MUS-50*, pages 45–50. IPSJ, 2003.
- [13] C. Raphael. Orchestra in a box: A system for real time musical accompaniment. In *IJCAI2003 workshop program, APP-5*, pages 5–10. IJCAI, 2003.
- [14] S. M. Shieber. Lessons from a restricted turing test. *Communications of ACM*, 37(6):70–78, 1994.
- [15] T. Taguti. <http://www.is.konan-u.ac.jp/tag-lab/mp3.htm>.
- [16] A. M. Turing. Computing machinery and intelligence. *Mind*, (236):433–460, 1950.