ASSESSING RHYTHM RECOGNITION SKILLS IN A MULTIMEDIA ENVIRONMENT [1]

Irene Cheng [2, 3], Chris Kerr[2] and Walter F. Bischof [2]

[2]Dept. of Computing Science, University of Alberta, Canada
[3]Dept. of Computer and Information Science, University of Pennsylvania, USA

ABSTRACT

Multimedia content is an effective tool for enhancing learning and testing, and it is more effective than the traditional paper and pencil presentation format. In order to make educational materials more intuitive, more interactive and more inspiring to students, image, audio, video, graphics, animation and 3D items are widely used in current educational applications. However, they are designed mainly for learning and not for testing. Those designed for testing focus on using simple, text-based formats (e.g. multiple-choice and True/False questions) to assess a student's subject knowledge rather than on evaluating a student's cognitive skills and problem-solving ability. In this paper, we propose an interactive framework that uses an innovative test format enriched with audios and videos, for evaluating a student's skills, in this case musical rhythms recognition skills. Experiments were conducted with human observers and the results verify the feasibility of our approach. The goal of this paper is to advance multimedia research in education by extending its capabilities beyond simple knowledge retention and reproduction, and to model the acquisition of cognitive skills in a dynamic context.

Index Terms— Cognitive Skill, Rhythm Recognition, Multimedia Test, Rank Coefficient, Interface Evaluation

1. INTRODUCTION

Musical intelligence is one of the seven intelligences defined by Psychologist Howard Gardner [1]. Among these seven intelligences, musical intelligence is very difficult to assess on paper. Musical intelligence is defined as the ability to perceive, discriminate, compose, express, transform and invent musical forms. An important component of musical intelligence is the rhythm recognition skill. In order to assess a student's rhythm recognition skill, one can proceed in several ways. Written exams can test the student's knowledge of music, including, for example, musical history. They can also be used to test a student's basic understanding of music rules. One disadvantage of written exams is a failure to capture the "live" spirit of music. In contrast, an interview-based test focuses more on a student's on-stage performance, including vocal and instrumental skill, but it has the disadvantage of requiring expert examiners.

The objective of our design is to introduce a computerized system that makes use of acoustic and visual effects for presenting an environment that can inspire the student's feeling of being engaged in a musical scene. Cross and Deliège [10] pointed out that there is a drive towards "computerization of music" − the embodiment of aspects of music in computer software and in hardware. The research on representing elements of music in computational terms was motivated by powerful aesthetic, educational and commercial imperatives. Inspired by this school of thoughts, we propose to use computerized test items to assess skills such as musical rhythm recognition.

Research on performance assessment has relied on the use of techniques such as Bayesian networks [2] or Machine Learning techniques [3], in which student models are based on the traditional goal of testing, i.e. knowledge retention. In contrast, our approach aims at testing a student's cognitive skills with innovative multimedia items.

Gardner [4] developed a number of ideas on how one can take a more scientific approach to the assessment of intelligences other than linguistic and logical-mathematical intelligence, which are tested in most standard evaluations. He stated "in order to test an individual's abilities in the bodily or musical domains, it is clearly inadequate simply to use paper-and-pencil. It should be possible to develop intrinsically compelling activities, e.g., simple games, which allow a ready assessment of an individual's interest, and potential for development in a given intellectual domain." [4]. While he did not go into the details of such assessments, he suggested that "the core of musical intelligence is rhythmic and pitch analysis" [4]. His suggestion fits well with Seashore's definitions in his book "The Psychology of Musical Talent" [5]. Motivated by their findings, we used an abbreviated version of the Seashore test to establish the

[1] This project is funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada; Informatics Circle of Research Excellence (iCORE), Alberta; Castle Rock Research Corp., Edmonton, Alberta; and Mathematics of Information Technology and Complex Systems (MITACS), Canada.

ground truth for our own experiments on the assessment of musical rhythm recognition.

The rest of this paper is organized as follows: In Section 2, we describe the preparation of multimedia test items and the method of calibration. Section 3 describes the experiments. Section 4 presents the results and analysis. Finally, Section 5 draws the conclusion and discusses future work.

2. PREPARATION OF MULTIMEDIA TEST ITEMS AND CALIBRATION

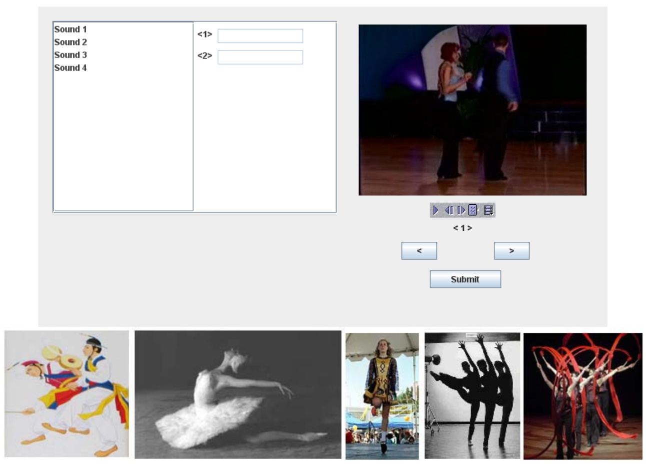

Figure 1: A screen shot of a multimedia item used in our experiment (top) and some examples of videos used (bottom).

Studies of the association between musical and film information revealed a direct influence of musical meaning on film meaning [11], suggesting that meaning at any point in time is the resultant of associations generated between the two media. Based on Cohen's finding, we used videos containing sequences of dancing patterns accompanied by sound tracks. Observers with better cognitive skills in recognizing musical rhythms should do better in associating an appropriate sound track to a dancing scene.

The goal of our research is to test the viability of multimedia based assessment of musical rhythm recognition skills, and thus a design of the test items is required. A preliminary template was taken from earlier work [6] and revised to meet the needs of our experiments (see Figure 1). The original template design required the observer to drag five out of seven audio clips from the left panel to the middle panel. The five boxes in the middle panel correspond to the five dance videos displayed sequentially in the right panel.

There are many factors that need to be taken into consideration when setting up audios and videos for a test item. For example, the test item should avoid displaying videos with significant cultural differences, which may provide obvious hints to the correct answer, even for observers with little musical knowledge. On the other hand, if all the audio and videos are similar, the difficulty level might be too high for the observers. Our goal was to obtain uniformly distributed scores on reasonably difficult test items. We therefore required a sufficiently large number of test items in our experiments. Due to the limited size of the test database, using too many clips in one item while achieving the above objective could be difficult. With a few observers, we tested using two videos and two audios. However, we found that the scores were high because the probability of guessing correctly was 50%. We then decided to match four musical clips with two dance videos without taking partially correct answers into consideration. An absolute score of either 0% or 100% was awarded for each test item. Statistically, the probability of guessing a correct answer was thus reduced to $(\frac{1}{4}\times\frac{1}{3}) = 8\%$.

2.1. Calibration of Test Items

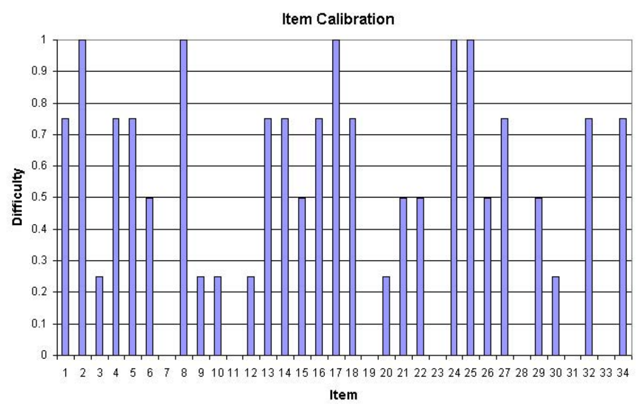

Figure 2: The calibrated difficulties of 34 test items.

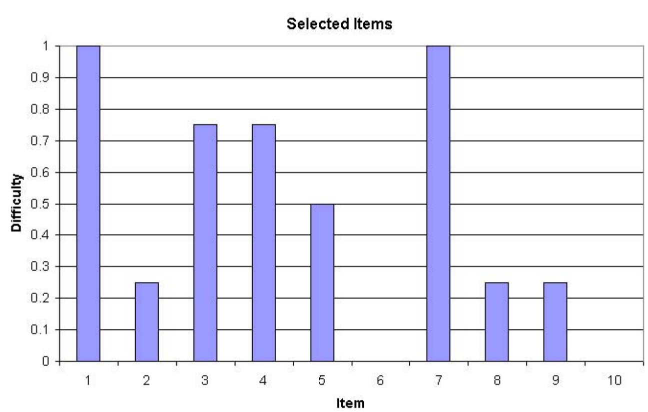

Figure 3: 10 calibrated items were selected with difficulties distributed in the range [0, 1].

71 videos associated with sound clips were carefully chosen to balance the requirements of diversity and similarity. The sound clips were then decoupled from the videos. 34 test items were composed from these elements, without reusing

the same video and audio in multiple test items because a student could otherwise rely on prior knowledge when he/she saw the video again.

In order to assign the difficulty levels to the test items, four observers with a broad spectrum of cultural, educational, and musical backgrounds participated in the calibration process. These observers were not involved in the final rhythm recognition skill evaluation. The difficulty of each item was computed as:

$$Difficulty = 1 - \frac{\#correct}{\#observers}, \tag{1}$$

where "#correct" is the number of observers who answered correctly and "#observers" is the number of observers who participated in the tests. The results for the 34 items are shown in Figure 2. From these results, we selected 10 test items representing a uniform distribution of difficulty levels (see Figure 3).

3. EXPERIMENTS

The goal of our experiments is to assess whether our multimedia framework can be used to accurately evaluate students' skills in musical rhythm recognition. Bigand [12] explored the issue of how a listener abstracts the elements of musical structure that may determine his/her emotional responses to music. By examining the interaction of sound and vision via film, his study shows how musical sound contributes to the observer's overall perception. We used dance videos instead of films to explore how sound tracks provide clues to the relation between musical structure and expression through body movements.

Procedure

The experiment was composed of three parts, which each observer had to complete. First, observers filled in a questionnaire to get a general impression of the observers' musical backgrounds. Second, an abbreviated Seashore [5] test was used to establish the ground truth of the observer's musical abilities. In this test, the observer must discriminate between two sounds, which may differ in pitch, intensity, or rhythm. Third, observers were tested with multimedia test item described in Section 2.

The observers performed the abbreviated Seashore test at the same time as they were completing the questionnaire. This was done in order to prevent the observers from actively memorizing of the sounds. The questionnaire contained simple questions about the observer's age, school year, and others. The observers were also asked to describe, in as much detail as possible, their musical background (e.g. attendance of music lessons, preferred music category, and frequency of listening to music). The observers also had to rate their own "musicality" on a scale from 1-10, with 10 indicating the highest musical skill. The answers were later compared to the results obtained from the multimedia test items.

The abbreviated Seashore test consisted of pairs of sound clips. There was a break of 45 seconds between each pair of clips. At the same time, observers were answering the questionnaire. This prevented the observers from actively rehearsing, forcing them to encode the sounds into memory and to recall them afterwards for comparison with the subsequent clip. Observers with greater musical aptitude were capable of encoding the information into memory more accurately [5]. A total of ten pairs of sound clips were used. These clips presented differences in pitch, tempo, note duration and rhythm. The observers were asked whether each pair of clips was the same, and, if not, how they differed (e.g. whether they were differed in duration, tempo, or pitch).

The calibration and the multimedia items were presented using the same computer. Each item was composed of four sound buttons and two sequences of dance videos (without music). The observer could double- click the sound description (sound 1 to 4) to play the music and had to match a video clip with one of the sound clips by judging the rhythm of the music and the dance movements. The observer had to drag the corresponding sound ids into the two answer boxes, one for each video, and then press the submit button to confirm the choice. The same procedure was repeated for the next item. The 10 calibrated items, selected from a pool of 34, as described in Section 2, were presented to the observers, and no feedback was given. The whole experiment, consisting of abbreviated Seashore test, the questionnaire, and the multimedia test took approximately 1 hour to complete.

Observers

Two groups of observers participated in the study. Group 1 consisted of 9 high school students (grades 10 and 11), and Group 2 consisted of 5 students in the age range 20-30.

Digital media

All video clips were 15 seconds long, had a resolution of 326x240 pixels, and a frame rate of 25 Hz.

Equipment

The experiments were performed on a 2.41 GHz AMD Athlon 64 processor with 1 GB RAM, with an NVIDIA GeForce 7950GT video card and a SoundMAX Integrated Digital HD audio card. The monitor size was 41 x 26 cm with a resolution of 1,440x900 pixels. The sound clips were presented over headphones.

4. RESULTS

We used the Kendall's Rank Correlation Coefficient τ [8] to describe the relationship between the ground truth, the performance in the abbreviated Seashore test and the performance with the multimedia test items. Let X denote the ranking of the n observers (n = 9 and 5 in Group 1 and

Group 2 respectively) in the abbreviated Seashore test and Y denote the corresponding ranking based on the multimedia test items. If there is a strong agreement in their ranking, we can conclude that our multimedia test items can accurately assess observers' rhythm recognition skills. Let (X_1, Y_1), … , (X_n, Y_n) be a bivariate random sample from the function F(X, Y), an absolute continuous distribution. Two observations (X_i, Y_i) and (X_j, Y_j) are concordant if both members of one pair exceed the corresponding members of the other pair. Concordance can be expressed as $(X_i, Y_i)(X_j, Y_j) > 0$. Thus the rank correlation coefficient τ is

$$\tau = \frac{4P}{n(n-1)} - 1, \quad (2)$$

where P is the number of concordant pairs.

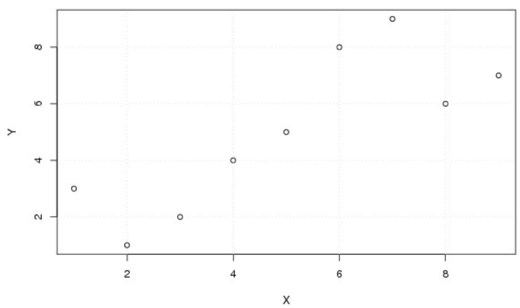

Figure 4: Rank correlation scatter plot for Group 1.

Rank coefficient τ	Group 1	Group 2
Multimedia test	0.67	0.80
Self Evaluation	0.11	0.00

Table 1: Rank coefficients.

For Group 1, the correlation coefficient was $\tau = 0.67$(Figure 4), which is significantly different from zero ($p < 0.02$). For Group 2, the correlation coefficient was $\tau = 0.80$ ($p < 0.08$). The discrepancy in the τ values can be explained by the maturity of the observers in the second group. Note that a similar situation can occur in any evaluation involving human observers, and a more controlled experimental setting is likely required for younger observers.

When checking the answers in the questionnaires, we noticed that self-evaluation of musical ability did not show a strong correlation with the ground truth. This can be explained by the Dunning-Kruger effect [7], which suggests that those with lower ability in a given area tend to over-estimate their skill levels due to insufficient understanding of the area's complexities. An examination of the correlation between the self-evaluation and ground truth ranking revealed $\tau = 0.11$ for Group 1 and $\tau = 0.00$ for Group 2 (Table 1). In contrast, the correlation between the multimedia test and ground truth was $\tau = 0.67$ and $\tau = 0.80$ respectively, indicating that our multimedia tests can assess observers' musical rhythm recognition skills better than self-evaluation.

5. CONCLUSIONS

In this paper, we studied the feasibility of multimedia tests for assessing musical rhythm recognition skills. Instead of using the traditional Seashore format, the current trend of educational testing is in computer-based and internet-based testing. By incorporating audio-visual effects, testing becomes more intuitive, more interactive and more effective. Computer-based testing reduces administration cost, can be performed at any time, can be scored automatically, and can be accessed remotely. We conducted experiments to verify our approach and found that the accuracy of our assessment on musical rhythm recognition skills is comparable to that obtained with the traditional Seashore test.

We believe that our result can be improved if a larger pool of test items and more observers is available. Furthermore, incorporating feedback from musical teachers may also improve the accuracy. Our future work will focus on analyzing the cultural background of observers and on using a different statistical model, Item Response Theory (IRT) [9], to further verify our multimedia framework.

6. REFERENCES

[1] H. Gardner, "Frames of Mind: The Theory of Multiple Intelligences," New York: Basic Books, 1983.

[2] Z.A. Pardos, N.T. Heffernan, B. Anderson and C.L. Heffernan, "Using Fine-Grained Skill Models to Fit Student Performance with Bayesian Networks", 8th International Conference on Intelligent Tutoring Systems, Jihongli, Taiwan, 2006.

[3] J.E. Beck and B.P. Woolf, "High-Level Student Modeling with Machine Learning," Proceedings of the 5th International Conference on Intelligent Tutoring Systems, 2000, pp. 584–593.

[4] H. Gardner, "Artistic Intelligences", Art Education, Vol. 36, 1983, pp. 47-49.

[5] C. Seashore, "The Psychology of Musical Talent," Boston: Silver Burdett Company, 1919.

[6] I. Cheng and W.F. Bischof, "Multimedia Item Type Design for Assessing Human Cognitive Skills", International Conference on Multimedia & Expo, Beijing, 2007.

[7] D. Dunning and J. Kruger, "Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments", Journal of Personal and Social Psychology, Vol. 77, 1999, pp. 1121-1134.

[8] M. Kendall, "A New Measure of Rank Correlation", *Biometrika*, Vol. 30, 1938, pp. 81-89.

[9] V. Linden, and R. Hambleton, "Handbook of Modern Item Response Theory", London, Springer Verlag, 1997.

[10] I. Cross, and I. Deliège, "Cognitive science and music – An overview", Contemporary Music Review, Vol 9, 1993, pp 1-6.

[11] A.J. Cohen, "Associationism and musical soundtrack phenomena", Contemporary Music Review, Vol 9, 1993, pp 163-178.

[12] E. Bigand, "The influence of implicit harmony, rhythm and musical training on the abstraction of tension-relaxation schemas in tonal musical phrases", Contemporary Music Review, Vol 9, 1993, pp 123-137.