RECOGNITION OF ISOLATED INSTRUMENT TONES BY CONSERVATORY STUDENTS

Asha Srinivasan, David Sullivan, and Ichiro Fujinaga

Computer Music Department

Peabody Conservatory of Music

Johns Hopkins University

Baltimore, MD USA 21202

ABSTRACT

An experiment was conducted in order to reconstruct previous timbre recognition experiments, measure the effect of ensemble experience and short-term training on the recognition rate, and generate more detailed baseline data to help evaluate the performance of timbre recognition computer models. The subjects, who were conservatory students, had to identify between 2, 3, 9, and 27 instruments on two different occasions: once without practice and once with short training sessions before the test. Eighty-eight subjects participated in the experiment. All tones were taken from the McGill University Master Samples CDs.

Compared to previous experiments, the average scores of subjects in this experiment were considerably higher. Additionally, subjects who play orchestral instruments scored significantly higher than those who do not. Finally, the short training sessions had no significant effect on the subjects' performance.

1. INTRODUCTION

Musicians have a remarkable ability to recognize instruments by timbre and, as both Eagleson & Eagleson (1947) and Martin (1998, p. 40) pointed out, it is a common perception that identifying musical instruments is easy. Little research, however, has been done on how well people can identify instruments. Some of this research has been conducted as experiments on various aspects of timbre, such as attack and release, and steady-state portions, on the effect of recognition rates under different environment conditions, and as a reference point for computer-based experiments. The information that can be gathered about human recognition rates from these experiments suggests that these rates are lower than the expected values. The experiments conducted using timbre recognition computer models have shown that these models can match or exceed the human rates.

The purpose of this experiment was to reconstruct previous timbre recognition experiments, measure the effect of ensemble experience and short-term training on the recognition rate, and generate more detailed baseline data to help evaluate the performance of timbre recognition computer models. The number of instruments in each section of this experiment correspond to that of four previous experiments: Brown (1999), Kendall (1986), Elliott (1975), and Martin (1998), for the 2, 3, 9, and 27-instrument sections, respectively. The main difference was that the subjects in this experiment were all conservatory students. It was predicted that conservatory students would do better in the tests than subjects from previous experiments. This would also provide music researchers with data on the recognition rates of the best human subjects.

2. METHOD

Participants: Eighty-eight subjects participated in the experiment. These included undergraduate and graduate ear-training students, composition students, and three faculty members from the Peabody Conservatory of Music.

Stimuli: All instrument samples were taken from the McGill University Master Samples CDs.

Equipment: The samples of the instruments were extracted from CD with a digital audio processing software. No editing was done to the samples. The digital audio software was simply used to locate the necessary pitches and then to save those as separate files. All of the instructional voice audio files were recorded using another digital audio processing software and then the gain was modified to match the levels of the instrument samples. All of the audio files, voice and instruments, were put together in a digital audio sequencing software, giving adequate spacing in between each question. The final sequence was made into one audio file, which was then burned onto a CD. The CD was played on a portable CD player with a woofer of 2.5" radius. The tests were presented in a classroom setting.

Procedure: For each test, isolated tones were played by the specific instruments in a section, presented in random order, and the subjects' task was to identify which instrument had produced each tone. Each tone was between 4-7 seconds long. Subjects were given 5 seconds to answer each question after the tone was played. They were given answer sheets with the instruments listed in multiple-choice format for the 2, 3, and 9-instrument sections and in word-bank format for the 27-instrument section. They were told to answer every question, even if they did not know the answer, thus employing the forced-choice method. Each test was burned onto a CD as one track with pre-recorded instructions, meaning that subjects could not ask any questions or stop the CD for any reason. This provided consistency for each time the test was given.

Two tests were performed. The first test included four sections, involving 2, 3, 9, and 27 instruments. The instruments are listed in Table 1. Before the first test, subjects had to fill out a personal information sheet, which asked about their musical experience.

The second test repeated the first test, except with short training sessions before each section. This test had the same answer sheet format and presentation style as the first test, but only three sections, involving 2, 9, and 27 instruments, were presented. Results from the first test showed that almost all subjects had a recognition rate of 95% or higher in the section with 3 instruments. Since there could not be any more significant improvement based on training in this section, it was abandoned for the second test. Also different on the second test were the pitches of the instruments and the order in which they were presented.

The short training sessions that preceded each section lasted 2 min., 3 min., and 10 min., respectively. In these training sessions, each instrument was identified by name before sounding and then several tones from each instrument were played. Then, the test for that section was given. In the case of the 27-instrument section, instruments were grouped based on similar sound and same family. A brief test followed each presentation of a group of instruments. This test consisted of four questions based on the instruments in that particular group. This was done mainly to maintain the subjects' engagement and did not effect the computing of the recognition rates. After all groups were presented, the subjects had to take the 27-instrument test again.

2-instrument	3-instrument	9-instrument
Oboe Sax	Clarinet Trumpet Violin	Flute Oboe Clarinet Bassoon Sax Trumpet Trombone Violin Cello

27-instr		
Violin	English horn	Tenor sax
Viola	Bassoon	Bari sax
Cello	Contrabassoon	Bass sax
Double bass	Eb clarinet	Trumpet
Piccolo	Bb clarinet	French horn
Flute	Bass clarinet	Alto trb
Alto flute	Contrabass clarinet	Tenor trb
Bass flute	Sop sax	Bass trb
Oboe	Alto sax	Tuba

Table 1: List of Instruments

3. RESULTS

Average scores in our tests were 94.5%, 97.6%, 90.2%, and 55.7% for 2, 3, 9, and 27 instruments, respectively (Fig 1). The recognition rates for the best student was 100% for the 2- and 3-instrument sections, 96.7% for the 9-instrument section, and 77.3% for the 27-instrument section of the first test. The student's rates for the second test were 100% on 2-, 3-, and 9-instrument sections, and 85.3% on the 27-instrument section.

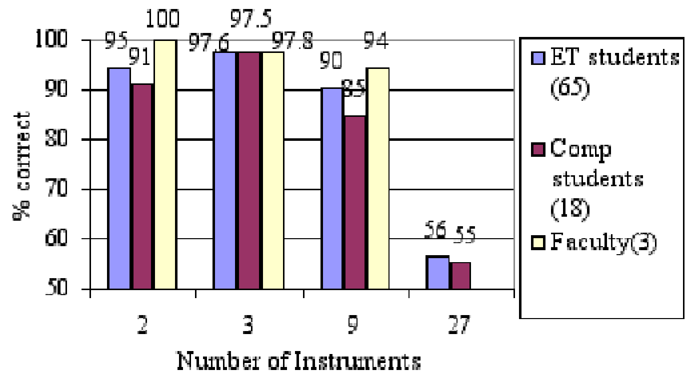

Figure 1: Recognition rates for ear-training (ET) students, composition students, and faculty.

Figure 2 shows our average rates in comparison to previous experiments: Saldanha (1964), Martin (1998), Eagleson & Eagleson (1947), Berger (1964), Elliott (1975), Strong & Clark (1967), Brown (1999), Kendall (1986).

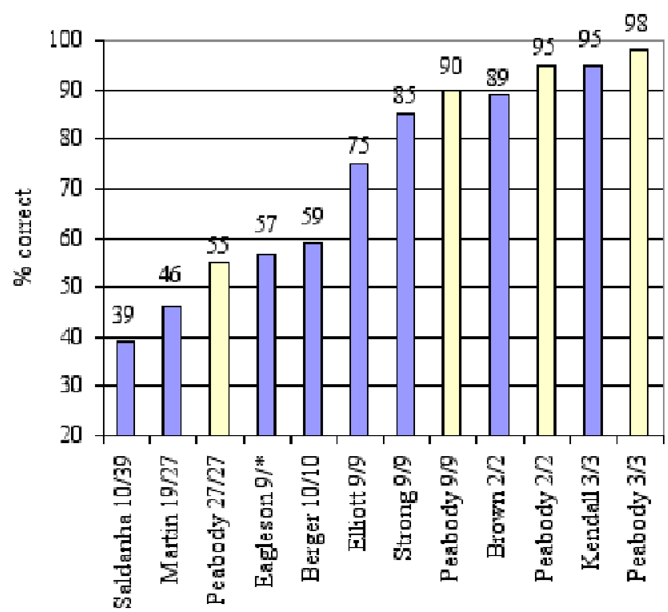

Figure 2: Recognition rates for previous experiments and Peabody's results. The number by the name of the experiment shows the number of instruments presented/the number of instruments given on the answer sheet. Eagleson did not give subjects a list of instruments, which is why there is an '*' after the '/'.

The recognition rates of piano, guitar, and voice (PGV) students were separated from the non-PGV students, as shown in Figure 3.

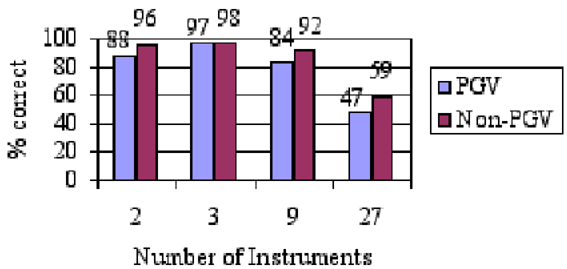

Figure 3: Rates of PGV vs. Non-PGV students.

Several confusion matrices were generated. Tables 2, 3, and 4 show confusion matrices for the 2-, 3-, and 9-instrument sections, respectively.

		1	2
1	Ob	95	5
2	Sax	4	96

Table 2: 2-instrument confusion matrix (87 subjects)

		1	2	3
1	Clar	98	2	
2	Trpt	4	96	
3	Vln			99

Table 3: 3-instrument confusion matrix (88)

		1	2	3	4	5	6	7	8	9
1	Fl	99								
2	Ob		92	3				3		
3	Cl		4	87	4			5		
4	Bsn		2	2	84		3	6		
5	Sax		2			97				
6	Trpt				5	8	83	2		
7	Trb			4				94		
8	Vln								84	17
9	Vc								9	91

Table 4: 9-instrument confusion matrix (88)

The 27-instrument confusion matrix was too large to display in this paper. Please contact the author to see it.

Answers were grouped into families to see whether this would improve recognition rates (Fig 4).

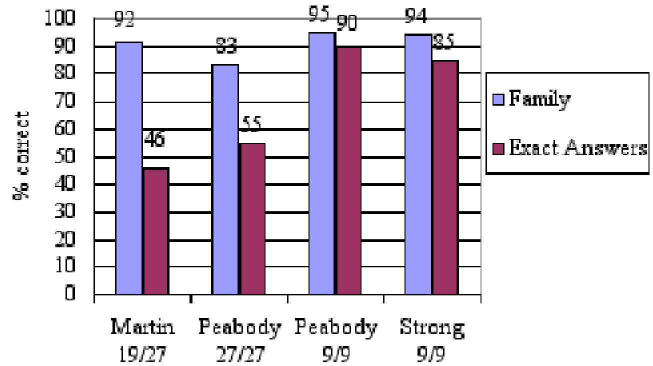

Figure 4: Family vs. Exact Answers

Finally, the results of second tests, taken after the short training sessions, were compared to the first tests, separately for the ear-training and composition students (Fig 5).

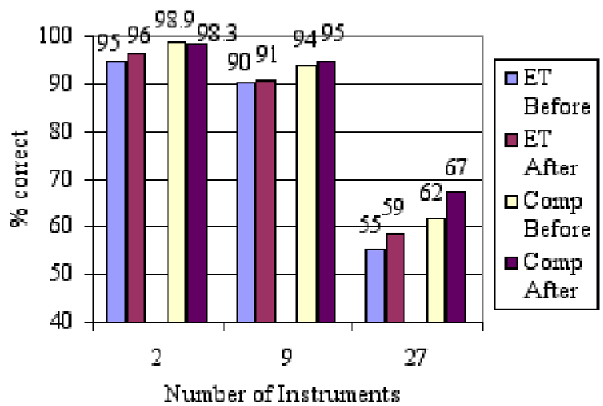

Figure 5: Effects of training on ear-training and composition students.

5. DISCUSSION

The average scores of subjects in this experiment were considerably higher than those of previous experiments (Fig 2). This could be because this experiment used conservatory students, however this cannot be said for sure. Using the personal information sheet, several t-tests were performed on different combinations to determine what was a significant factor in higher or lower recognition rates. In doing so, it was found that subjects who played orchestral instruments had significantly higher recognition rates than subjects who were pianists, guitarists, or singers (Fig 3). This could be why the ear-training students did slightly better than the composition students (Fig 1) because most of the composition students were pianists or did not have much orchestral experience.

The confusion matrices show that there was relatively little confusion in the 2-instrument and 3-instrument sections (Tables 2 & 3). Additionally, in the 3-instrument section, the violin was rarely confused with the other two instruments. A few people, however, did confuse clarinets and trumpets.

The 9-instrument confusion matrix is a little more revealing. As can be seen in Table 4, violin and cello were confused for each other, but not with any of the other instruments. Flute was also not confused with the other instruments. The 27-instrument confusion matrix shows that there was confusion between the woodwind and brass instruments. There was also confusion between the lower instruments, such as contrabassoon, contrabass clarinet and bass sax. There is a strong similarity between the confusion matrices of this experiment and other experiments, for both the 9-instrument, as compared to Strong's matrix (1967), and the 27-instrument sections, as compared to Martin's matrix (1998, p.143).

The short training sessions did not have a significant effect on the recognition rate for most of the sections. Although in the 27-instrument section, there did seem to be a small improvement between the first test and the second test, there was no significant improvement in recognition rates between the first and second test, even when answers were grouped into families.

6. CONCLUSIONS

Compared to previous experiments, the average scores of subjects in this experiment were considerably higher. This experiment showed that musicians who play orchestral instruments are better at recognizing timbre than musicians who play instruments that are not usually found in the orchestra, such as piano, voice, and guitar. The short training sessions did not have a significant effect on most sections of the test. There was a slight improvement, however, in the 27-instrument section. Overall, this experiment presents new challenges to computer timbre recognition models by giving researchers better baseline data to rely on for human recognition rates.

7. REFERENCES

1. Berger, K. W. "Some factors in the recognition of timbre", *J. Acoust. Soc. Amer., Vol. 36, 1964, p.1888-1891.*

2. Brown, Judith. "Computer identification of musical instruments using pattern recognition with cepstral coefficients as features", *J. Acoust. Soc. Amer., Vol. 105 (3), 1999, p.1933-1941.*

3. Corso, J.F., and Saldanha, E. L. "Timbre cues and the identification of musical instruments", *J. Acoust. Soc. Amer., Vol. 36, 1964, p. 2021-2026.*

4. Eagleson, H. V., and Eagleson, O. W. "Identification of musical instruments when heard directly and over a public-address system." *Journal of the Acoustical Society of America, Vol. 19 (2), 1947, p. 338-342.*

5. Elliott, Charles. "Attacks and releases as factors in instrument identification", *J. Research in Music Education, Vol. 23, 1975, p. 35-40*.

6. Kendall, Roger A. "The role of acoustic signal partitions in listener categorization of musical phrases", *Music Perception, Vol. 4 (2), 1986, p.185-214.*

7. Martin, Keith D. *Sound-Source Recognition: A Theory and Computational Model.* Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1999.

8. Strong, W. & Clark, M. "Synthesis of wind-instrument tones", *J. Acoust. Soc. Amer., Vol. 41 (1), 1967, p.3*