Malaysian Journal of Computer Science, Vol. 12 No.2, December 1999, pp. 64--70

COMPUTER MUSIC PLAYBACK QUALITY: DIGITAL AUDIO REPRODUCTION VERSUS SYNTHESIZED SOUND

Minni Ang[1] and Sook Sian Wong[2]
Music Department
Faculty of Human Ecology
Universiti Putra Malaysia
43400 UPM Serdang
email: [1]minni@music.upm.edu.my
[2]sooksian@yahoo.com

ABSTRACT

The music we hear played back via computer controlled media is generated by one of two technologies: digital audio, which is a waveform reproduction of the original sound source recorded and played back through digital means, and musical instrument digital interface or MIDI, which communicates a set of performance instructions to a synthesiser chip which generates playback in real-time. The ability to distinguish synthesized sound from digital reproductions of original sound sources is tested through a set of listening tests conducted on music technology undergraduate students. Research findings indicate that certain synthesizers and soundcards are able to produce high quality synthesized sound that is perceived to be as good as the original instrument timbres.

Keywords: ***Computer Music, MIDI, Synthesisers, Digital Music, Soundcards, Music Synthesis, Music Playback Quality, Digital Audio***

1.0 INTRODUCTION

Recent advances in improving the quality of music playback have proceeded along two separate fronts. Digital audio reproduction has progressed to the accompaniment of blaring media headlines, as sampling rates and bit resolutions hit unprecedented highs. Coupled with exponential improvements in the capacity of storage media, such as DVD, the enormous additional space required for such high quality audio files has not proved a hindrance.

On a quieter note, advances in music synthesis technology have also been significantly improving the quality of music output through synthesizers, which may be found as dedicated hardware or as an integrated computer soundcard component [1, 2].

To the user who is unaware of the differences in the technologies involved, digital audio and MIDI file playback may appear to be merely different file formats with different extensions, all of which are easily played back through the modern personal computer.

The digital audio file stores the actual waveform of sounds within a series of binary words obtained through the sampling process. The playback of such a file is therefore original and realistic, with clarity depending only on the sampling rate and bit resolution of the stored binary words. The MIDI file stores a set of performance instructions, such as what note is to be played by what instrument and when, but does not specify how the individual instrument sounds should be produced [3].

The issue studied in this research is whether or not users are truly able to distinguish between the audio output produced by these two vastly different technologies. MIDI files, which contain only performance instructions, are extremely small when compared with digital audio files of the same duration. The advantage of using MIDI as opposed to digital audio is therefore immediately apparent to those who would include music playback in their computer-based applications. MIDI file playback is, however, dependent on the playback device used to interpret these performance instructions. Traditionally, this has meant that applications using MIDI have produced inconsistent playback quality, with even the best playback devices sounding slightly artificial in nature. A side point to note here is that songs using the human voice can never be accurately played back via MIDI, as performance instructions cannot include pronunciations of words. However, current developments in music synthesis technology have resulted in highly realistic simulations of most musical instruments. If it can be shown that MIDI files can produce consistent and high quality playback indistinguishable from digital audio recordings of the same music, then computer game programmers, web masters and other people who program music within computer applications can use this much more compact medium with confidence that their product will sound the way it is supposed to sound.

2.0 LITERATURE REVIEW

Dedicated hardware synthesizers, synthesiser chips in computer soundcards, and software synthesizers are the three major implementations of music synthesis technology which are utilised to playback MIDI files. Various synthesis techniques are used, the most frequently found on commercial synthesizers striving to imitate real-life instruments being wavetable or sampling synthesis [4], which creates musical tones from recorded samples of actual acoustic or electric instruments. Older soundcards and budget models often use frequency modulation (FM) synthesis [4, 5]. The latest high-end synthesizers utilise physical modelling which provides even greater control over various performance parameters [6, 7].

Research on the comparative sound quality of playback devices has been carried out for various computer soundcards [8]. However, tests have focused only on quantitative parameters such as frequency response, signal-to-noise ratio, dynamic range and jitter, and have not taken into account users' preferences. Other studies conducted [9] have indicated that frequency response, for example, seems to correlate with results of subjective listening tests, but specific subjective listening tests with the aim of discovering users' ability to distinguish between different playback devices were not carried out.

3.0 METHODOLOGY

MIDI files with instrumentation and arrangement identical to original musical excerpts available on audio compact disk were selected for comparison, with short excerpts of 14 to 16 seconds being chosen for this purpose. Each MIDI file was played back using different synthesiser devices and the output was recorded, in random sequence, onto compact disk alongside the original instruments recording. 45 undergraduate students majoring in Music Technology then participated in a series of listening tests, in which individual preferences towards the excerpts were assessed through a questionnaire, on a scale of 1 to 10, with 10 representing the best liked sound quality. Participants, who were not informed as to the exact number of original instrument recordings which were included in each test, were also requested to decide whether the excerpt was from a MIDI source or was an original recording. The sample population chosen for this test was for the deliberate intention of enhancing the validity of the results obtained - the test candidates being all highly trained musicians familiar with MIDI and digital audio. Other sample populations, for example the general public or musicians not familiar with the differences between MIDI and digital audio, may perhaps not fully understand the difference between digitally generated music and digitally recorded music, and hence be unable to ascertain how a particular excerpt may have been produced. Another point to note is that test candidates pre-dispositions towards either medium have no bearing on the results obtained, as they were not informed as to which excerpts were from which source.

Besides the original audio CDs, 15 MIDI playback devices were utilised. This selection was based on the availability of devices at the institution where the research was conducted, and included 5 computer soundcards, 3 hardware synthesizers and 5 software synthesizers (Table 1). The different devices used different synthesis techniques, including FM and wavetable synthesis, and physical modelling (Table 1).

Table 1: MIDI Playback Devices Used for the Test

Synth Type	***Device***	***Basic Synthesis Method Utilised in Tests***
Soundcard	SB AWE 32	MIDI Synth: wavetable
	SB AWE 64	MIDI Synth: wavetable MusicSynth: FM
	SB Live!	MIDI Synth A: wavetable MIDI Synth B: wavetable
	Acer Magic	FM
	ESS Tech	ES688: FM
Software	Roland Virtual SC55	wavetable
	Yamaha SXG50	wavetable
	Yamaha OPL3 (installed with SB AWE32 card)	FM
	Creative SoftSynth (installed with SB Live! card)	wavetable
	Creative WaveSynth (installed with SB AWE64 card)	wavetable
Hardware	Yamaha EX5	wavetable, physical modelling
	Casio CTK-750	wavetable
	Roland SC55	wavetable

Table 2: Musical Excerpts Utilised in the Test

No.	*Title of Piece*	*Artist / Composer*	*Genre*
1	Beat It	Michael Jackson	Rock
2	Etude in G Flat	Chopin	Solo Piano
3	The Spanish Flea	Herb Alpert	Latin
4	Dance of the Sugar Plum Fairy	Tchaikovsky	Orchestral

A total of four musical excerpts were selected for comparison: the opening bars of a famous rock song, a classical piano solo, a Latin brass number, and a symphony orchestra excerpt (Table 2). These musical excerpts were selected to represent different major music genres. The listening test is thus comprised of four musical excerpts played back 16 times each, totalling 64 clips of music. To avoid listener fatigue and thus less validity of the test results, short breaks were allowed between tests, with excerpts played back as often as requested by the test candidates. Participants were also required to decide, for each of the four musical excerpts, which of the 16 playback devices they most preferred.

3.1 Analysis Methods

The analysis of results was carried out in two stages. Stage 1 involved an excerpt by excerpt analysis of the data collected. This included calculating the mean value and standard deviation values of individual preferences towards particular playback devices, calculating the percentage of respondents who thought that a particular device played back a MIDI file or an audio file, and ranking the most preferred sound through calculating the percentage of respondents who selected a particular device for this role. Stage 2 of the analysis involved the calculating of an overall ranking of sound choices based on mean values of rankings for individual musical excerpts, and rankings of the most preferred sound for individual musical excerpts.

4.0 RESULTS

4.1 Analysis of Musical Excerpt No. 1

The analysis of data collected for musical excerpt No. 1, "Beat It", by Michael Jackson, showed that most respondents preferred the original audio CD recording, which obtained (on the scale of 1 to 10) a mean preference value of 8.13. The second highest ranked clip was the Yamaha SXG50 software synthesiser, which obtained a mean preference value of only 6.24. This result shows that the respondents were clearly able to distinguish between the superior sound quality of the original recording, as opposed to the best quality synthesized sound. It is noted here that the excerpt selected included only musical instrument sounds and not the singer's voice because, as pointed out earlier in this paper, it is impossible to use MIDI to reproduce pronunciations of words. Three sound devices obtained mean preference values of less than 4 points - the AWE32 Creative OPL3, the AWE64 Creative Music Synth, and the Acer Magic soundcard. All these three devices use FM synthesis. This result seems to clearly indicate that FM synthesis is of inferior sound quality to other synthesis methods. Most respondents were able to accurately identify the original audio CD clip as an original recording, and the other clips as MIDI files. 65 percent of respondents chose the original audio CD clip as their favourite sound, with small percentages (between 2 to 8 percent each) selecting other clips as their most preferred sound. While this indicates the majority prefers the original sound, it is also significant to note that a significant minority actually prefers the synthesised clips (which are identified as such) to the original. Table 3 and Fig. 1 below provide a summary overview of the results obtained for Musical Excerpt No. 1.

Table 3: Top and Bottom 3 Rankings Derived from Mean Preference Values for Musical Excerpt No. 1

#	*Sound Device*	*Mean Preference*	*Standard Deviation*
1	Original CD	8.133	1.809
2	SXG50	6.240	1.714
3	SBLive! Synth B	5.711	2.062
14	OPL3	3.822	1.690
15	AWE64 MusicSynth	3.640	1.791
16	Acer Magic	3.330	2.458

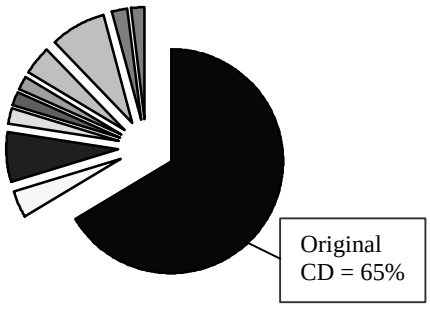

Fig. 1: Most Preferred Sound for Musical Excerpt No. 1

4.2 Analysis of Musical Excerpt No. 2

Analysis of the data collected for musical excerpt No. 2, "Etude in G Flat" by Chopin, produced very interesting results. In terms of mean preference values as well as most preferred sound, the Yamaha EX5 hardware synthesiser came out in top rank, with ratings of 6.91 and 21% respectively. The top four ranking mean preference values, all over 6 points, all fell to synthesized sounds, with the original audio CD recording only gaining seventh placing at 5.62 points. Interestingly, the original sound also had the largest standard deviation value for mean preference at 2.52, which indicates a relatively large inconsistency of preference towards this sound by respondents. In the most preferred sound rankings, the Yamaha SXG50 and the AWE64 MIDI Synth fell a close second and third, with 18% and 16% of the vote respectively. Most respondents incorrectly identified the top six ranked sound devices as being original audio recordings, but also correctly identified the actual original recording as an original recording. Again, the bottom four rankings fell to devices using FM synthesis, and all were correctly identified as MIDI files, with mean values all below 3.4 points. Table 4 and Fig. 2 below provide a summary overview of the results obtained for Musical Excerpt No. 2.

Table 4: Top and Bottom 3 Rankings Derived from Mean Preference Values for Musical Excerpt No. 2

#	*Sound Device*	*Mean Preference*	*Standard Deviation*
1	EX5	6.911	2.042
2	AWE64 MIDI Synth	6.067	1.960
3	CTK-750	6.044	2.108
14	OPL3	3.133	1.984
15	Acer Magic	3.089	2.020
16	AWE64 MusicSynth	2.533	1.962

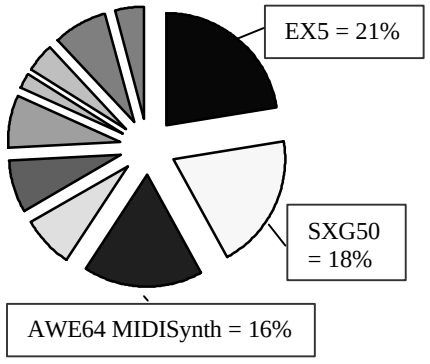

Fig. 2: Most Preferred Sound for Musical Excerpt No. 2

4.3 Analysis of Musical Excerpt No. 3

Musical excerpt No. 3, "The Spanish Flea", by Herb Alpert and the Tijuana Brass, produced inconsistent results. Based on mean preference values, the Casio CTK-750 hardware synthesiser slightly outranked the original audio CD recording at 7.62 points to 7.51. However, 39% of respondents selected the original recording as their favourite sound, compared with 34% who chose the Casio sound. Most respondents thought both clips were original audio recordings. Many respondents were also confused with the clips produced by the Yamaha EX5 and Roland SC55 hardware synthesizers, with more than 40% of respondents in each case stating that the clips were audio files, with another group of more than 40% in each case stating that these same clips were MIDI files. Both these devices did not fare so well in the most preferred sound rankings though, gaining only 4% and 13% of the vote respectively. Once more, the bottom four rankings fell to devices using FM synthesis, and all were correctly identified as MIDI files, with mean values all below 4.3 points. Table 5 and Fig. 3 below provide a summary overview of the results obtained for Musical Excerpt No. 3.

Table 5: Top and Bottom 3 Rankings Derived from Mean Preference Values for Musical Excerpt No. 3

#	*Sound Device*	*Mean Preference*	*Standard Deviation*
1	CTK-750	7.622	2.120
1	CTK-750	7.622	2.120
2	Original CD	7.511	2.051
3	EX5	6.756	2.120
14	AWE64 MusicSynth	4.067	2.091
15	Acer Magic	3.956	2.319
16	OPL3	3.889	2.213

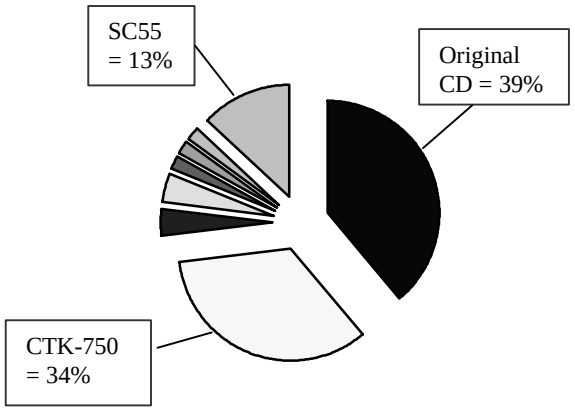

Fig. 3: Most Preferred Sound for Musical Excerpt No. 3

4.4 Analysis of Musical Excerpt No. 4

The final set of data, for Musical excerpt No. 4, "Dance of the Sugar Plum Fairy" by Tchaikovsky, produced similar results to that for Musical excerpt No. 2. Here, the Yamaha EX5 hardware synthesiser again gained top rank, both in terms of mean preference values and in most preferred sound rankings, at 7.33 points and 52% respectively (Table 6, Fig. 4). The results this time indicated a clear lead, with the second mean preference value being 0.7 points less, and the second ranked most preferred sound gaining only 11% of the vote. The original audio CD recording only gained fifth place in the mean value rankings, and only 11% of the most preferred sound vote. Most respondents thought that the EX5 clip was an original audio recording. Although the original audio recording itself only gained fifth place, most respondents also correctly identified it as an original recording, while at the same time also accurately identifying the second through fourth placed mean preference value rankings as belonging to MIDI files. Only the EX5 clip was incorrectly identified in this respect. The last four places were once again filled by devices using FM synthesis, this time with mean values all below 4.4 points.

Table 6: Top and Bottom 3 Rankings Derived from Mean Preference Values for Musical Excerpt No. 4

#	*Sound Device*	*Mean Preference*	*Standard Deviation*
1	EX5	7.333	2.181
2	AWE64 WaveSynth	6.644	1.84
3	SXG50	5.844	1.897
14	ESS ES688	4.333	2.119
15	Acer Magic	3.489	2.353
16	AWE64 MusicSynth	3.378	1.936

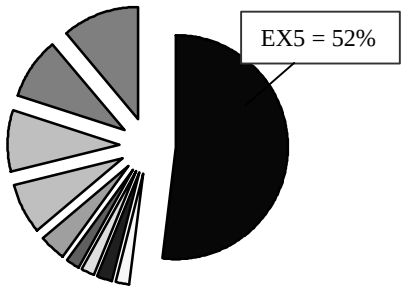

Fig. 4: Most Preferred Sound for Musical Excerpt No. 4

Table 6 and Fig. 4 above provide a summary overview of the results obtained for Musical Excerpt No. 4.

4.5 Overall Rankings

Table 7 below displays the overall ranking of all playback devices (including the original audio CD recordings) based on rankings obtained through mean preference values. The values in this table were calculated from the averages of the rankings obtained through mean preference values for each of the four musical excerpts, and range from values of 1 through 16, with lower values indicating a better ranking.

Table 7: Overall Rankings Derived from Mean Preference Value Rankings for Individual Excerpts

Rank	*Sound Device*	*Mean Value of Rankings for Individual Musical Excerpts*
1	Yamaha EX5	3.00
2	Original CDs	3.75
3	Casio CTK-750	4.00
4	Yamaha SXG50	4.25
5	AWE64 Wave Synth	6.25
6	SB Live! Synth B	6.25
7	Roland SC55	7.00
8	SB Live! Soft Synth	7.25
9	AWE64 MIDI Synth	8.00
10	Roland VSC55	8.25
11	SB Live! Synth A	8.75
12	AWE32 MIDI Synth	11.25
13	ESS Tech ES688	13.25
14	AWE32 OPL3	14.25
15	Acer Magic	15.25
16	AWE64 Music Synth	15.25

In this table, the Yamaha EX5 hardware synthesiser emerges as the top ranked playback device, better than the original recordings which are ranked second. The Casio CTK-750 hardware synthesiser and the Yamaha SXG50 software synthesiser come a close third and fourth.

In terms of most preferred sound, overall percentages are calculated based on average percentage values for each of the four musical excerpts. In this case, the original CD recording gained overall top rank, with the Yamaha EX5 synthesiser gaining one in five votes, and the Casio CTK-750 one in ten. Table 8 displays the top eight ranking sounds. The other eight devices scored less than these values, and account for less than 10 percent of all choices.

Table 8: Overall Rankings Derived from Most Preferred Sound for Individual Musical Excerpts

Rank	*Device*	*Percentage*
1	Original CD recording	30
2	Yamaha EX5	20
3	Casio CTK-750	11
4	Yamaha SXG50	8
5	AWE64 MIDI Synth	6
6	Roland SC55	6
7	SB Live! Soft Synth	5
8	AWE64 Wave Synth	4

5.0 DISCUSSION

The results for Musical Excerpt No. 1 indicate that the sound of the electric guitar (contained in the musical excerpt) is difficult to synthesise convincingly, as most respondents were able to identify the original from the synthesized sounds. In spite of this, a significant minority still preferred the synthesized sounds to the original, perhaps indicating in future that high quality electronically generated sound may one day become an accepted norm. The qualifier "high quality" is used here, as in all the tests, the sounds produced through the undeniably artificial sounding FM synthesis devices were always the least preferred sounds. These results further imply that, while FM synthesis may have its place in the world of synthesized music, it cannot be used to convincingly simulate real-life instruments. As such, soundcards and other MIDI playback devices using this synthesis technique, will always be perceived as inferior when playing back commonly found General MIDI (GM) files, which usually try to imitate real-life acoustic instruments.

It should be considered that the quality of the original audio recordings for Musical Excerpt No. 2 and No. 4 might have affected the judgement of respondents. Both were commercially available original audio CD recordings of professional musicians of top calibre - however it is possible that the recordings were not of sufficiently high quality, and when played back against crystal clear MIDI generated files, suffered in comparison. Results for Musical excerpt No. 4 provide further confirmation that respondents sometimes prefer high quality synthesized sounds over lower quality original recordings, as the second through fourth most preferred excerpts were all correctly identified as MIDI files and yet still preferred over the correctly identified original audio recording, which gained only fifth place. Having said that, the results also clearly indicate that the sound quality produced by the best MIDI playback devices could not be identified as not being produced by acoustic instruments. It is clear that piano sounds, and certain orchestral instrument sounds, are able to be reproduced accurately by modern synthesizers, especially those utilising wavetable or sampling synthesis technology.

The results for Musical Excerpt No. 3 further confirm the confusion respondents had in distinguishing between the best quality synthesized sounds and digital audio reproductions. The overall rankings show that the best quality synthesized sounds, produced by hardware synthesizers, are on par with original recordings in terms of sound quality. The best performing software synthesiser puts up a respectable showing at fourth place, though clearly behind the leaders in terms of preference by respondents. The best performing soundcard, from the five that were tested "out of the box", that is without any tweaking or installing of sound fonts or other extra sound banks, was the Sound Blaster AWE 64. These results may be surprising, considering that better quality sound cards were also tested - however, better quality sound cards have greater expansion possibilities than the budget models, and may perhaps perform better on these tests if all options were fully utilised.

6.0 CONCLUSIONS AND SUGGESTIONS FOR FURTHER STUDY

The primary conclusion, which can be drawn from this research is that generally, top quality synthesizers produce sound outputs, which are very difficult to distinguish from the original sounds. The state of current synthesis technology is undoubtedly very good at synthesising solo piano sounds, and also very good at synthesising orchestral instrument timbres when played in ensemble. However, not all instrument sounds can be imitated flawlessly. In the tests conducted during the course of this research, it is seen that it is difficult to accurately simulate the sound of the solo electric guitar. From here it may be summarized that the sound of solo string instruments are generally hard to imitate.

A further conclusion, which may be drawn, is that, in certain instances, synthesized sounds may actually be preferred to original acoustic instrument timbres. Previous studies have shown that listeners do not generally prefer electronically generated sounds [10]. The results of the present research may reflect a change in perceptions towards electronically generated sounds, by a generation of young people who have grown up accustomed to these sounds.

Other conclusions are that dedicated hardware synthesizers produce the best quality of sound output, though some software synthesizers perform almost as well as hardware synthesizers. Quality of sound produced by sampling synthesis soundcards has improved tremendously compared with older FM synthesis type cards, but still do

not measure up to the standards set by the best hardware synthesizers.

Finally, FM synthesis never produces good enough sound quality to match original instrument timbres, and neither does it produce pleasing enough sounds for it to be preferred over other sound sources. As such, for sound card manufacturers to remain competitive, emphasis should be on producing sound cards using wavetable or sampling synthesis technology.

Further study needs to be done to determine which specific instrument sounds can be accurately imitated by various synthesis methods and why. Comparison of quality of sound output needs to be done for a greater variety of soundcards, with various options installed. This would help in determining the universality of playback of MIDI files by computers, especially in the context of MIDI files embedded into various applications. Consistency of high quality sounds for MIDI playback across soundcards would enable software developers to use this technology, rather than the bulkier digital audio file formats, for integrating music into various applications.

REFERENCES

[1] Stanford University News Service, "Music Synthesis Approaches Sound Quality of Real Instruments" in *Stanford News*, 1994. Available: *http://www.stanford.edu/dept/news/relaged/940607Arc4222.html.*

[2] M. Karjalainen, V. Valimaki, and Z. Janosy, "Towards High Quality Sound Synthesis of Guitar and String Instruments", *Proceedings of the 1993 International Computer Music Conference*: San Francisco, 1993, pp. 56-63.

[3] C. Dodge and T. Jerse, *Computer Music: Synthesis, Composition and Performance.* New York, Schirmer Books, 1997.

[4] J. Heckroth, *Tutorial on MIDI and Music Synthesis.* La Habra CA, The MIDI Manufacturers Association, 1995. Available: *http://www.harmony-central.com/MIDI/Doc/tutorial.html.*

[5] F. R. Moore, "Dreams of Computer Music – Then and Now." *Computer Music Journal*. Vol. 20 No. 1, 1997, pp. 25-41.

[6] D. Nomer, "Physical Modelling", 1998. Available: *http://web800.com/music/vl/physmodl.htm.*

[7] X. Rodet, "Recent Developments in Computer Sound Analysis and Synthesis". *Computer Music Journal*. Vol. 20 No. 1, 1997, pp. 57-61.

[8] A. Krueger, "PC AVTech Sound Card Technical Benchmarks Test Summary", 1999. Available: *http://www.pcavtech.com/soundcards/summary/index.htm*.

[9] Y. Tannaka, "Correlations Between Sound Field Characteristics and Subjective Ratings on Reproduced Music Sound Quality". *The Journal of the Acoustical Society of America.* Vol. 85 No. 2, August 1989, pp. 603-620.

[10] M. Bridger, "Narrativization in Electroacoustic and Computer Music: Reflections on Empirical Research into Listeners' Response". *Proceedings of the 1993 International Computer Music Conference.* San Francisco, International Computer Music Association, 1993, pp. 296-299.

ACKNOWLEDGEMENT

This work was made possible by facilities provided at Universiti Putra Malaysia.

BIOGRAPHY

Minni Ang received her B.Sc.Hons. in Physics from Universiti Malaya, her GBSM in Percussion Performance from Birmingham Conservatoire, and her Ph.D. in Computer Applications in Music from Universiti Putra Malaysia. She is currently the Head of the Music Department at UPM. Her current research interests include music synthesis technology and musical instrument acoustics. She is a member of the Audio Engineering Society, the Acoustical Society of America, the American Institute of Physics, and the International Computer Music Association.

Sook Sian Wong received her B.Mus.Hons. in Music Technology from Universiti Putra Malaysia. She is currently employed as a Systems Analyst with Hitachi Consumer Products (M) Berhad. Her research interests are in the area of computer based hardware for applications in music technology.