

Concurrent Auditory Stream Discrimination in Auditory Graphing

Hong Jun Song and Kirsty Beilharz

Abstract—This paper is concerned with enhancing human computer interaction and communication in concurrent streams of auditory display. Auditory display or auditory graphing is the sonic representation of numerical data (the auditory equivalent of visualization). It provides an additional channel for information representation, in which a participant's response may be more intuitive and immediate than (visual) graphical display, but auditory graph design requires understanding and multi-disciplinary investigation of listening due to instantaneous characteristics of sound. Our aims are to explore (1) the impact of spatial separation for a divided attention task and (2) the efficiency of timbre (tone color) to assist pitch contour identification. Our findings about timbral and spatial discrimination are scalable and useful for auditory display in a wide variety of contexts. The results provide empirical evidence for a further investigation of spatialization and timbre and contribute to applications within an auditory display context for real-world scenarios (e.g. social, statistical and other datasets likely to be encountered in the workplace).

Keywords— Auditory graph, spatial separation, timbre, divided attention, concurrent streams.

I. INTRODUCTION

A. Background

SONIFICATION delivers information through non-speech sound [1]. Due to its intuitive connection with symbolic meanings, sound has been used extensively as a means for Human Computer Interaction (HCI), where it can provide rapid comprehension of the processing status, alarms or warnings related to a system. With the remarkable growth of digitization in our society, auditory display meets the needs of extending computational methods for communicating and analyzing data, where sound is able to furnish immediate feedback on our actions.

Auditory graphs or sonification are broadly defined as a means to display quantitative data (e.g. scientific data or real-time financial data) [2]. In a time-series application, the y-axis of auditory graphs are composed by pitches mapping from the numerical data values of graphical lines and the x-axis corresponds to the movement of the pitch over time. Visualization of complex data normally has dense overlapped graphical lines or clusters. The use of sonification to display these quantitative data has the potential to relieve visual overload. However, the principles for auditory graphing are not as well documented as for visual graphing and designing a

good auditory graph requires effective methods of representing data variables (mapping). When working with multi-variate quantitative data, difficulties arise because of (a) the interaction of auditory parameters (e.g. pitch, loudness and rhythm) that can influence one another when several auditory parameters are used to represent data variables, (b) interference of audio streams (i.e. masking) when simultaneous auditory streams are applied to represent different data dimensions, and (c) multi-tasks (e.g. divided attention) when multiple auditory stream patterns require simultaneous deciphering or recognition.

B. Multi-dimensional auditory streams display

Concurrent presentation increases information display dimensions and allows parallel processing [3][4][5]. Depending on the numbers of auditory streams being recognized, the tasks are compartmentalized into divided and selective attention. In divided attention situations, both of the concurrent signals need to be identified; in selective attention situations, only a single important audio stream needs to be extracted from competing auditory messages [6][7][8]. The known principles of listeners' ability to separate individual strands of meanings from sound mixture and grouping overall meanings are incorporated in the perceptual framework of Auditory Scene Analysis (ASA) [9].

The ultimate goal of sonification design is to display various types of information clearly and unambiguously. Representing multi-dimensional data through auditory display is extremely challenging because complexity arises with multiple audio streams, in which audio streams reach to our ears as a mixture. It is easy to cause confusion and blending of auditory streams. Thus mapping strategies have to meet both the requirements of "informativeness" (of quantitative data) and "simplicity/clarity" (for the listener).

C. Aim

This study aims to examine the design choices for perceptually promoting auditory grouping ability in auditory graphing and to contribute to the body of empirical evidence. We discuss the potential benefits of spatial separation and timbre (tone color) on concurrent stream representation and provide results in a task of pitch contour recognition in order to improve sonification design strategies.

Spatialization and timbre are the two salient auditory attributes facilitating a better comprehension of sonification when simultaneous audio streams are delivered. While directional cues provide obvious interpretable structure for auditory display, timbre provides semantic expression by highlighting identity of auditory stream. In this paper, the

Manuscript received October 9, 2006; Revised version received March 4, 2007.

H.J. Song and K.A. Beilharz are at Key Centre of Design Computing and Cognition, Faculty of Architecture, Design and Planning, University of Sydney, Sydney, NSW, Australia (e-mail: hson5790@usyd.edu.au; kirsty@arch.usyd.edu.au).

primary investigation is how sound dimensions (spatialization and timbre) could be designed to achieve a distinguishable display of the information contained in auditory graphs and how effectively these dimensions could be used for better comprehension.

II. SPATIALIZATION FOR STREAM SEGREGATION

While spatial separation is not as frequently discussed as pitch among the available auditory parameters, we advocate that it is an effective dimension for auditory display [10]. Spatial separation has a better “force and semantic structure” than pitch to reduce “problems of peripheral sensory masking” [11] and maintain attentions to the directions of sound sources. Directional cues, acting as a spotlight, enhance the processing of sounds by combining spectra-temporal features of binaural signals and speed up discrimination responses by providing an essential interpretive context that gives distinctive structure to sound. On the basis of experience, directional cues aid differentiation of subjective mental representation of audio streams [12]. The benefits are from two features of spatialization. Firstly, spatialization is relatively independent and bears no common forms with other parameters. Thus spatial representation makes audio streams reasoning “easy”. The fundamental frequency and location of a tone can be changed with negligible interaction, whereas if fundamental frequency and sound pressure level are changed independently, ambiguity may occur due to the interaction between these dimensions (rising in pitch also causing perceptual increasing in loudness). Secondly, interaural time/level differences reduce masking and interference between streams [13] and improve the signal-to-masker ratio and the audibility of the target that allow listeners to direct their attention to the target.

In the study of Lorho et al.[10], the performance of an absolute spatial discrimination task shows that both stereo panning and head-related transfer function (HRTF) presentation over headphones allow an efficient spatial separation for a limited number of sound sources. Another study about earcon identification has shown that spatially located concurrent earcons were more easily identified than those having unique location [14]. Furthermore, research in speech field consistently illustrates the benefits of spatial separation for splitting competing verbal messages displayed either by headphones or by loudspeakers [13] [15][16]. One famous phenomenon is cocktail party effect which is related to speech intelligence and speaker recognition, referring to the ability of listeners to separate a single talker from competing talkers and background noises and to concentrate on specific conversation [17]. However, while intuition indicates that for *non-speech* sound spatial separation potentially is able to enhance auditory stream segregation, there is little empirical literature supporting and investigating spatial effect. Some have argued that the effectiveness of spatial separation is restricted to categorical data [17] and in divided attention situations the temporal information about individual strands of two auditory graphs will be destroyed due to individual’s spatial cognitive ability [18]. The findings within visual research [19][20][21] (which is usually analogous to auditory

display) also demonstrate that when spatial tasks were involved, there was a decrease of efficiency and with the augment of the spatial items in the search array, reaction time to the visual tasks increased and accuracy were reduced.

Traditionally, these arguments have caused some hesitation in using separated spatial locations for competing audio streams in sonification. When two sound sources are oriented from different directions, the listener needs to keep tracking and integrating auditory messages from two spatially separated locations. Spatial rehearsal allows tracking the orientation locations in working memory and spatial reasoning is augmented other than masking and interference of two audio streams. Therefore, divided attention, involving both competing auditory messages and imposing additional cognitive demands, probably decreases performance [22]. This study is motivated by the need to systematically explore and demonstrate the ways in which binaural spatialization influences auditory stream segregation.

III. TIMBRE FOR BETTER PERFORMANCE

The algorithms for sound generation (digital waveforms) are usually implemented by computer programs. These synthesized sounds are designed with certain characteristics by varying sound attributes. Timbre, known in psychoacoustics as sound quality or tone color, often is used to refer to all sound attributes that are not loudness, pitch and tempo. Timbre is defined as “an attribute of auditory sensation, in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar” [23]. Concerning features of timbre, many studies illustrate the important role of timbre for stream segregation, such as ASA. Furthermore, timbre plays a crucial role for an effective auditory display. It is essential to determine the acoustic variables (e.g. harmonics/inharmonics content) to produce convincing synthesised timbre for the stimuli during the mapping process, aiding interpretation of sonification content. By measuring the perceptual response to the generated sounds and characterizing timbral effect, the creation of desired sounds are deliberated concerned with comprehensibility and sound quality. One of the previous studies done by Bonebright [24] explored the perceptual space of a large set of everyday sounds and created three sound families by using the “multidimensional scaling solution” analysis. His three groups of sounds were defined according to correlations of acoustic measurement (e.g. average/peak intensity) and attributes rating 74 everyday sounds (a comparison between pairs of characteristics, e.g. dull/sharp). Bonebright’s results aim to benefit designers determining synthesis algorithms for sound generation to obtain better performance of auditory display.

Our concern in this study is to generate dissimilar timbres with an FM synthesizer for different pairs of audio streams in order to reinforce spatialized display. The timbral quality of fourteen synthesis sounds are investigated. The two main properties related to timbre (attack time and spectral centroid) were evaluated integrated with the participants’ performance and a correlation was found between characteristics of timbre and correlation index of pairs of graphs.

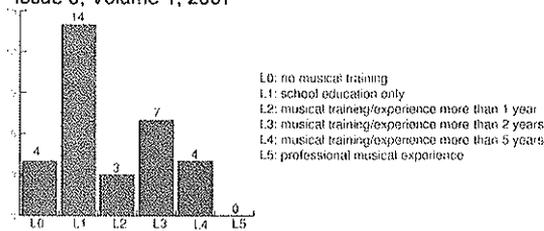


Fig. 1. Pie chart of musical background (MBG). Six levels were rated from beginner to expert. Numbers of people at each level are marked in corresponding pie segments. No participants had professional musical training, so the pie chart does not include this segment.

IV. METHOD

While both spatial location and timbre can yield improvement in pitch contour recognition by reducing interference between audio streams, this paper will focus on exposing the effect of directional separation to control divided attention (identifying bi-variate data streams) and discovering the potential physical characteristics of timbre augment or reinforce and enhance perception.

Participants tasks is to identify the contours of two concurrent auditory graphs when the sound sources were co-located and spatially separated. In this study, HRTF was used for binaural presentation to create the sensation of virtual direction of sound sources. The effect of spatial directional separation is evaluated by comparing the performance in co-located and spatially separated display modes. The current study quantified the benefit of spatial directional separation of auditory graphs for divided attention task in which source locations were fixed from trial to trial.

A. Participants

Altogether, thirty-four volunteers were recruited for this study. Most of them were enrolled students in the "Sound Design and Sonification" class at University of Sydney but they had not yet been exposed to discussion on spatialization at the time of the test. Data from two participants has been excluded: one due to an accidental equipment problem; and another due to excessive deviation from the group mean (this participant reported not reading the instructions carefully). The selected participants ranged in age from 19 to 41 years with a mean age of 23 years and there were 15 females and 17 males. All volunteers provided their informed consent according to the Ethics Committee of the University.

In the post-questionnaire, their musical background (MBG) was rated by themselves at six levels from beginner to expert (Figure 1). Most participants are at level 1 (14 of 32), in which they have school education only. Four people stated that they have more than two years of music training. None had professional musical experience.

B. Stimuli

The auditory stimuli were created in Max/MSP [25]. Max/MSP is a real time graphical programming environment. The parameters such as playing tempo can be easily altered

TABLE I
Stream configuration.

Parameters	Within a pair		Between pairs	Between modes
	Stream A	Stream B	Stream A&B	Stream A&B
pitch	changing [92, 587Hz]	changing [830, 1567Hz]	-	-
amplitude	constant	constant	-	-
timbre	constant	constant	changing [A, N]	-
virtual location 0/+60 degree	0/+60 degree	0/-60 degree	-	changing (0 or +/-60)
virtual distance	constant	constant	-	-

by editing the graphic icons. Its combination with visual display is useful for training or demonstrating audio outcomes for participants after the experiment. The SPAT library for Max/MSP [26] supports the HRTF function for binaural synthesis, so it is a spatialization library in which artificial reverberation, localization of sound source and spatial content of the room effect are integrated in a single processor patcher. It allows flexible and precise control of these effects [26].

The experiment used headphones for binaural reproduction with generic HRTFs. No head-tracking was involved. While binaural presentation technology could cause two problems: cone-of-confusion errors (such as front-back confusion) and head-locking of the sound-field (the sound-field moves with the listener's head), the spatial separation for the stimuli in this study was lateral and so relies predominantly on binaural difference cues, which are conveyed effectively using this technology. Spatial rendering was simply in terms of image direction, without any attempt to vary distance or other aspects of auditory space (e.g. vertical variation).

The configuration of auditory graphs within and between streams, pairs and modes is illustrated in Table I. The two auditory graphs correspond to the auditory streams (Stream A and B) respectively. Within one pair of streams, they differed in pitch register; different pairs had its own timbre; virtual locations of audio streams varied from a single position to two separated positions.

Taking account of ubiquitous and profitable listening to sound, the industry potential of sonification includes reinforcing visualization in bi-modal display situations and providing a non-visual alternative for embedding information efficiently. For this reason, binaural display and non-individual HRTFs are used (for isolation and affordability) that mimics the workplace environment where we envisage this sonification occurring. We recognize that the accuracy of head-tracking, individualized HRTFs or loudspeaker reproduction is greater but these laboratory conditions do not simulate workplaces, such as stock trading data analysis context.

1) *Mapping auditory graphs*: Twenty pairs of graphs were selected for sonification, and the concurrent audio streams are represented in Figure 2. Pairs of graphs are combinations of simple curve lines, straight linear lines or steps. The correlation between pairs was measured to

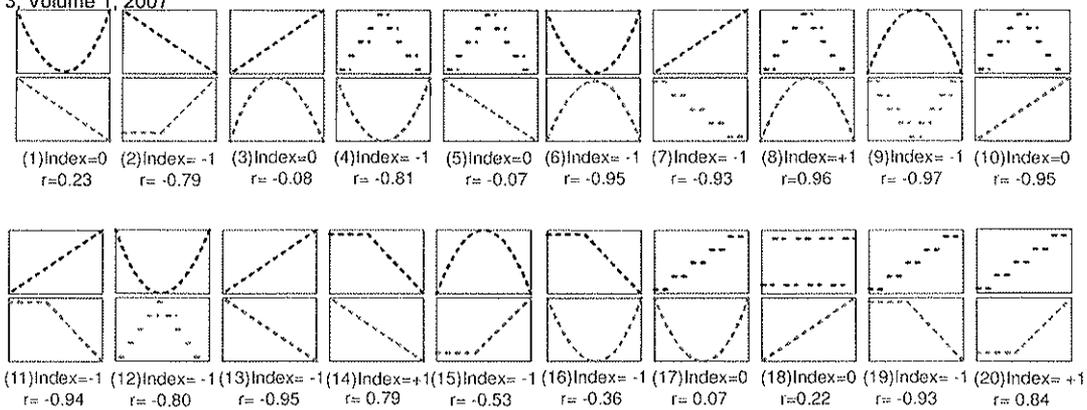


Fig. 2. Twenty pairs of graphs for sonification numbered from (1) to (20). Values of each graph were mapped onto pitches and each graph represents one audio stream. The index shows the correlation between two auditory graphs where “-1”: negatively correlated, “+1”: positively correlated and “0”: no correlation. The index was generated according to the Pearson correlation coefficients (r).

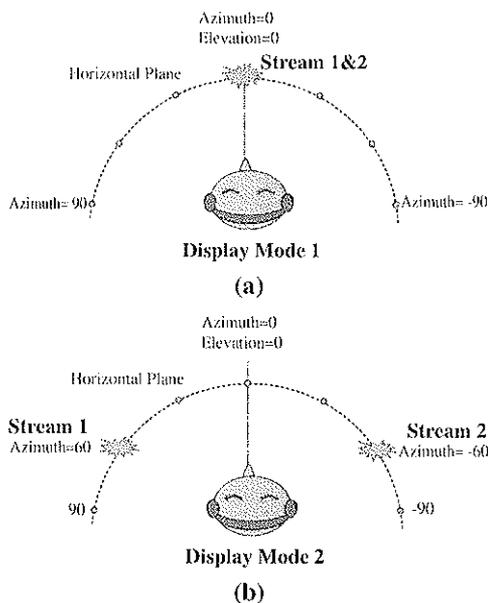


Fig. 3. Spatial configuration. Mode 1: both streams are from the middle line (azimuth= 0° and elevation= 0°); Mode 2: they are spatially separated, 60° degrees away from the middle line symmetrically.

explore the dependence of co-variation and coincidence between graphical lines and performance (percentage of errors occurred). The pairs of lines were marked as “-1”, “0” and “+1”, corresponding to negative, null and positive correlation respectively. The values of graphs were mapped onto MIDI notes ranging from 42 to 91, in which the x-axis of the graph was mapped onto time and y-axis of graphical lines was mapped onto midi note value. The speed was set to 10 points/second and the duration of each stimulus was about 10 seconds. The MIDI note was then converted to frequency in a logarithmic scale in order to be used for oscillator. Pitch register was applied to separate the audio streams. For the two concurrent pitch contours, one was located in pitch range from 92 to 587 Hz and another was from 830 to 1567 Hz.

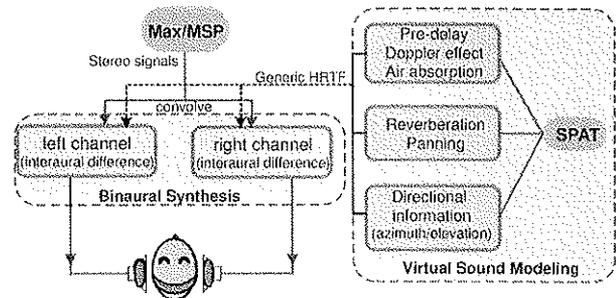


Fig. 4. System Overview. The two core units consist of virtual sound modeling and binaural synthesis, which render the virtual realism and realistic sound. Pre-delay (Doppler Effect, air absorption and reverberation) and pre-equalization are accomplished by SPAT. Generic HRTF are used to enhance binaural presentation.

2) *Reproduction of virtual spatial location:* Each pair was played twice in the two different display modes (Figure 3). In the single sound source condition, both of the concurrent audio streams are from the same sound source, which is 0° azimuth (straight ahead) in the horizontal plane (elevation 0°); or they were spatially separated. Best et al. found that interference still occurred up to 60° separation of verbal messages [16]. According to the previous findings [23][27] for concurrent minimum audible angle (CMAA), our case used 120° separation between two concurrent sound sources (60° to middle line). In order to generate sources those were perceived at different virtual spatial locations, the two auditory graph signals were convolved with anechoic head-related impulse responses (HRIRs) measured on a manikin head (KEMAR) for sources for either 0° (straight ahead) or $+60^\circ$ to the left of the middle line and -60° to the right of the middle line (all at 0° elevation and 1m distance). The HRTFs codes were collected by MIT’s Media Laboratory [28]. Interaural level differences (ILD) were added to the output channels according to the positional information of the sources. The cues of virtual sounds such as position, direction, distance, orientation and room effect were controlled with SPAT (Figure 4).

3) *Timbre generation with FM/AM synthesis* : Fourteen timbres were generated applying for different pairs of auditory graphs (Table II). Frequency modulation (FM) relies on modulating frequency of a carrier waveform with another modulator waveform. FM synthesis for this experiment was carried out by using simple FM (one resonance) or complex synthesis (two resonances) in which the two modulators were implemented in series. The linear envelope was shaped by amplitude modulation (AM) with alteration of graphical breakpoints and straight lines, and with various relative durations in Max/MSP.

Feature extraction of timbre is based on spectral centroid $f(c)$. Spectral centroid corresponds to the brightness/sharpness of a sound, which is perceived brighter with more high frequency components. There are other techniques such as tristimulus [29] (which describes timbre equivalent to colour attributes in vision) and spectral irregularity/smoothness [30][31] (in which the average of current, next and previous amplitude is compared with current amplitude). We use spectral centroid to judge timbre space is because it is based on physical property of a tone rather than cognitive recognition or non-acoustical grouping scheme.

There are other techniques such as tristimulus (which describes timbre equivalent to colour attributes in vision) [29] and spectral irregularity/smoothness (in which the average of current, next and previous amplitude is compared with current amplitude) [30][31]. The spectral centroid corresponds to brightness/sharpness of a sound which is perceived brighter with more high frequency components.

TABLE II

Configuration of timbres. Attack duration (ms) includes attack, sustain, decay and release; and last column are the graph pair ID corresponding to Figure 2.

Timbre ID	Spectral Centroid (Hz)	Attack duration (ms)	Applied graph pairs
A	1363	4000	(1) (13)
B	2659	1000	(2)
C	780	1000	(3) (19)
D	1601	160	(4)
E	5885	5000	(5) (11)
F	1767	200	(6)
G	1871	5000	(7) (17)
H	1137	5000	(8) (14)
I	721	2000	(15) (20)
J	1042	400	(10)
K	3001	4000	(12)
L	2677	400	(16)
M	1264	2000	(18)
N	1090	400	(9)

C. Task and procedure

The tasks were to listen to the graph sonification containing a pair of simultaneous audio streams and match the contour of both audio streams with the visual representations of Figure 2. Those binaural stimuli were presented with a CD player using six pairs of dynamic open-air Sennheiser HD 433 headphones. Those twenty pairs of graphs were displayed as co-located sound sources and as spatially separated sound sources so altogether there were 40 concurrent audio displays. They were

arranged in three random sequences to lessen the effect of sequence. Participants were divided into three groups of almost equal size, corresponding to one of the three random sound sequences. The purpose of the experiment was not mentioned in order to avoid listeners being distracted by seeking spatial cues. Instead, before starting the experiment, participants were provided with a one-page instruction sheet that included basic information about the experiment such as tasks and the duration. All participants received equal information from the paper instruction. Since participants pointed out that they became familiar with the display of the stimuli and felt comfortable after first a few trials, the first five trials were excluded from analysis and regarded as training. For each trial, the auditory stimuli consisted of a pair of contours (10 s) and a repetition (a 7.5s pause between individuals). After the repetition, listeners were required to circle one correct answer from 12 options in 7 seconds. Each trial started with a male voice announcing the question number and there was a beep prompt tone before the replay.

After the experiment, participants were required to fill in a questionnaire, in which they were asked to self-evaluate their musical background (MBG), indicating whether they noticed the different spatial display and providing basic personal information.

V. RESULTS

By comparing the difference of performance when the pair of audio streams were spatially separated or co-located, the overall performance showed that spatial separation **does** improve listener's ability to attend to two competing pitch contours when using binaural representation. As the purpose of this study was not mentioned before the experiment, most listeners stated in the post-experiment questionnaire that they were not aware of the spatial separation and were not conscious of the change of the spatial cues. In the survey, a few participants stated that they perceived the difference of spatial display mode but intuitively they thought in spatially separated display the concurrent contours were more easily recognized than co-located. The performance of 32 participants is summarized in Figure 5.

A. Discriminability of spatial separation

According to their performance, the thirty-two subjects are grouped into 4 categories: better, same, worse or no wrong answer. If the wrong answers in a single sound source (midline) exceed those for spatially separated sound sources (equally spaced along the azimuth), the sample belongs to category of "better", which means spatial separation enables better discrimination between two concurrent audio streams; if the wrong answers in the two display modes are the same in number, the sample is classified as "same", which means the discrimination is not evident; if the wrong answers in the single mode are fewer than in the separated display, the sample is classified into "worse", which means spatial separation did not enhance deciphering; people who did not make any wrong answers (total wrong answers = 0) are in group of "No wrong answer". Group behavioural data are summarized in Table III.

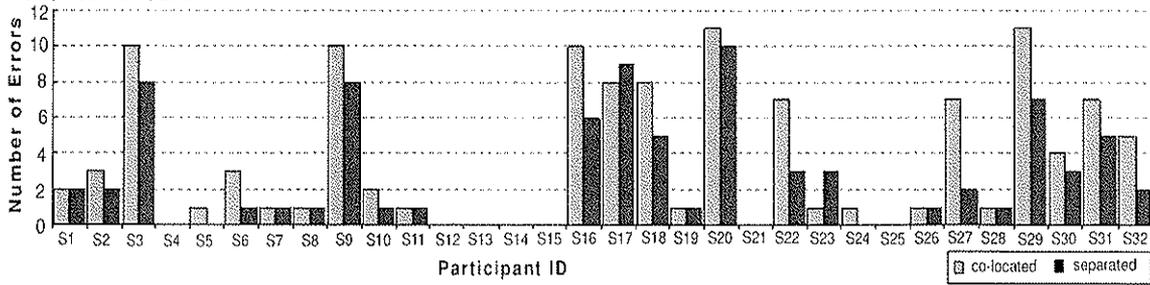


Fig. 5. Summary of occurred errors in co-located and separated modes. Subjects with ID S4, S12, S13, S14, S15, S21 and S25 had made no errors in both of the display modes; S5 and S24 only have errors occurred in co-located mode.

TABLE III

The performance when corresponding to two concurrent audio streams.

	Frequency	Percent
“better” ¹	16	50.0
“same” ²	7	21.9
“worse” ³	2	6.3
“no wrong answer” ⁴	7	21.9

- ¹ Errors (mode 1: single sound source) > Errors (mode 2: spatially separated sound source)
- ² Errors (mode 1: single sound source) = Errors (mode 2: spatially separated sound source)
- ³ Errors (mode 1: single sound source) < Errors (mode 2: spatially separated sound source)
- ⁴ Errors (mode 1) = Errors (mode 2) = 0

The two groups of wrong answers (each group for one display mode) are dependant and each set of paired wrong answers is from the same sample/subject. A paired t-test compared each set of pairs and analysed a list of difference between two groups. The result $[t(31)=3.968, p=0.005<0.01]$ demonstrates a significant difference between the two display modes. One sample t-test $[t(31)=3.968, p<0.01]$ showed that the difference of population between the “better” group and other groups is significant, which means performance is better when sounds were emitted from two fixed separated locations than from a single location. The result confirms the utility of spatial separation for concurrent audio stream in divided tasks and it indicates that spatial separation can be used in mapping to distinguish simultaneous concurrent data streams in information sonification when monitoring competing information streams is required.

B. Influence of graph correlation

Figure 6 plots the results for divided attention tasks as a function of graphical correlation index (corresponding to Table I). (a) Errors dropped when two signals were spatially separated (dashed line). The deviation is more obvious when they are not correlated (“0”) than negative (“-1”) and positive correlation (“+1”). (b) Fewest errors occurred when the two graphical lines were varied to the same direction (“+1”), and most errors occurred when there was no correlation (“0”) between two graphical lines. When the two signals were co-located (solid line), listeners have more difficulties identifying non-correlated lines (“0”). Additionally, it was

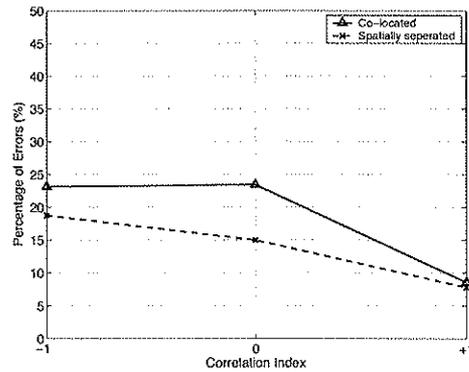


Fig. 6. The results are shown for co-located and separated sound source location (solid and dashed lines respectively), which plots the mean errors in recognizing the pairs of auditory graphs with graphical correlation index. It demonstrates that fewer errors occurred when sound sources of the two concurrent auditory graphs were from separated locations (bottom dashed line) whatever the correlation of two graphical lines was. The correlation index indicates the co-variation relationship of the two graphical graphs (shown in Figure 2).

found that the auditory graph pairs combined with steps (e.g. graph No.4 in Figure 2) were more easily identified. Since different performances were found between pairs with steps (26.1% errors) and without (36.8% errors), steps (modal pitch quantization) to represent data will be employed in future sonification applications.

C. Timbral quality ranking

Timbre in this paper is considered as a dependent parameter and the analysis of its impact was confounded with the pitch contours. According to total errors occurred in both of the display modes, the ranking of timbres is shown in Figure 7. The bottom three timbres are H, I and K, which have fewer errors than others and the top three timbres are D, F and L, which performance is worse than other.

Concerned the ranking of average errors for each timbre in Figure 7, Figure 8 demonstrates the extracted features of the three best and worst performing timbres. The results are the time dependent distribution of the centroid of the signal. The common features of top three is short attack time and overlapped spectrum while bottom three has long attack time but spectra are not overlapped.

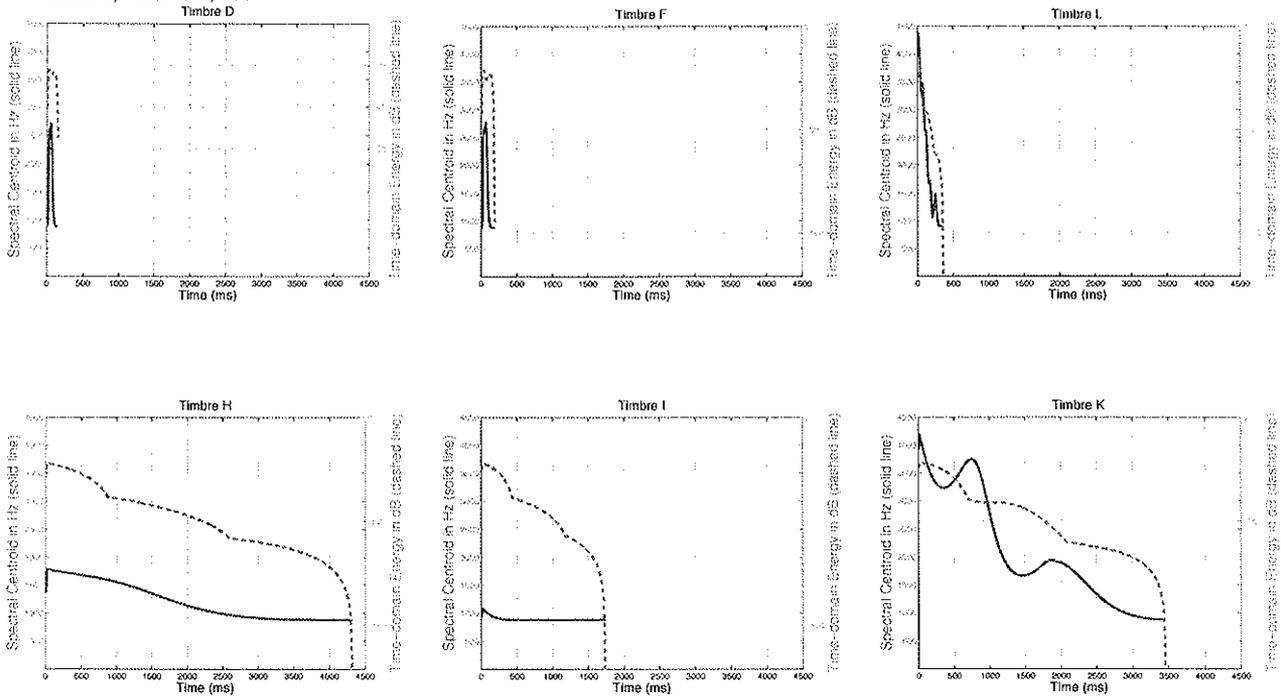


Fig. 8. Feature extraction of an 880Hz tone with the most effective and least effective timbres, where FFT size=1024 (Blackman windowing) and hop-size=512. The attack periods (attack, sustain, decay and release) of Timbre D, F and L were shorter than H, I and K.

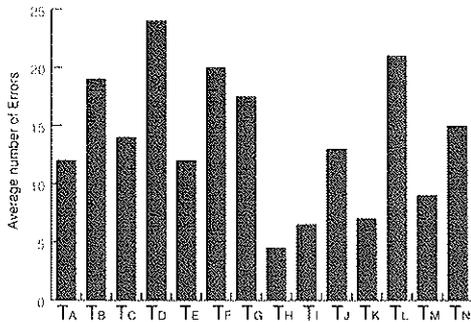


Fig. 7. Average errors of 14 timbres in both spatial display modes. Timbre D, F and L produce the highest number of errors, while timbre H, I and K have the fewest errors.

VI. DISCUSSION

This section examines the influence of two human factors on the performance (individual musical experience and gender). The results illustrate that they have no impact in the pattern recognition tasks. It provides evidence to determine effective strategies in training session.

It is uncertain what listening strategy participants were using in this experiment (divided or selective attention) because the stimuli were repeated once. Our intent was to reduce the complexity of the tasks but this might make it possible that listener probably is aware of the layout of two audio streams and then concentrate on one stream at the first time and on the other at the second time, especially after several questions.

A. Does musical background influence listener's performance?

Musical experience was considered as one of the human factors (such as age and gender) related to cognitive ability in many previous studies. In Neuhoff et al's case [32], musical experts and musical novices responded differently when the tasks was pitch magnitude estimation. Prior musical knowledge and expertise contributed to reaction time and the accuracy of interpreting pitch magnitude. The differences between expert and non-expert occurred in such area like memory, selective attention and categorization and they suggested that "if frequency change is to be used as a dimension to represent a variable in a display, then changes in frequency employed should be sufficiently large in order to minimize errors in judging the direction of change." Such findings have not been consistent being not sure the role of other individual cognitive difference factors in auditory interpretation. By replicating the Neuhoff et al's approach of assessing musical experience, for the tasks of auditory graph interpretation, Walker and Mauney indicated that musical background is not a significant contributor [33]. We were interested to discover if there was any correlation between musical knowledge and interpretation of auditory graphs with the augment of binaural cues, as we speculated that people with high level of musical training would make fewer errors.

One-way ANOVA [$F(4,27)=0.352 > 0.05$] has found no significance among six levels. This result indicates that musical background does not significantly influence their performance. Then group 0, 1, and 2 were re-grouped as low-level musical experience, and 3 and 4 as relative high-level

group. The comparison of the mean of these two groups [$t(30)=1.447$, $p(0.158)>0.05$] also shows that the level of musical experience does not affect the pattern recognition. This result is consistent with the findings of Walker and Mauncy [33], which indicates that musical background does not always influence on their performance rather depending on the tasks.

B. Does gender influence their performance?

There are almost equal numbers of male and female samples (15 female and 17 male). The factors of gender and age are often examined in relation to their impact on performance. Both male and female groups have samples which made no wrong answers (6 and 2 people). The result of ANOVA suggested that there was no difference in performance between female and male listeners [$F(1,30)=0.090>0.05$].

VII. FUTURE WORK

The results clarify the effect of directional cues facilitating concurrent auditory graph identification. However, the study of timbre is coarse and the decision of timbre in this study is based on a rough selection from a small set of FM/AM synthesised tone colors with primary consideration of dissimilarity. Out next work will focus on the evaluation of synthesised timbres which aims to find out applicable synthesised timbres from a listener's perspective by integrating perceptual characteristics of generated tone colors with acoustic measurement. The synthesis will be implemented by introducing multiple carriers either in series or parallel. The benefit of this is that spectral shape can be controlled independently. Granular synthesis technique is considered to be included to create a steady state and soft attack and decay. The evaluation is designed to incorporate perceptual characteristics of tone color with acoustic measurement.

Future work also will apply our findings in a sonification context. The potential application includes finding an effective way to explore or monitor data by using auditory display. Combining with other communication modes such as visual display, bimodal display will be considered but auditory display will be our priority focus. The representation system of complex data sets (large and highly dimensional) will try to satisfy aesthetical and functional requirements. The purpose of the proposed sonification system is to continue to qualify the benefit of spatialization and desired timbre in a sonification context. In the future if complex tasks (e.g. monitoring multi-stream large scientific data sets) are involved, there is an increasing demand for deliberate training strategies for users. We assume that training on specific tasks will decrease the difficulties of the monitoring or exploratory tasks and improve the performance, according to the previous studies on human capability of auditory learning and adaptation [34][35][36]. Based on the role of training and lack of musical experience of participants, the future experiment will apportion greater time to training. To encourage participants to follow the sound, the idea of including slight interaction will be considered, i.e. participants can trigger a human computer interaction (HCI) when they find any primary feature.

REFERENCES

- [1] G. Kramer, "An introduction to auditory display," in *Auditory Display: Sonification, Audification and Auditory Interfaces*, Gregory Kramer, Ed. 1994, vol. XVIII, pp. 1-77, Addison-Wesley.
- [2] M. A. Nees and B. N. Walker, "Relative intensity of auditory context for auditory graph design," in *12th International Conference on Auditory Display (ICAD'06)*, London, UK, 2006.
- [3] C. Hayward, "Listening to the earth sing," in *Auditory Display: Sonification, Audification and Auditory Interfaces*, Gregory Kramer, Ed. 1994, vol. XVIII, pp. 369 - 404, Addison-Wesley.
- [4] R. Minghim and A. R. Forrest, "An illustrated analysis of sonification for scientific visualization," in *Visualization 95*. 1995, pp. 110-117, IEEE.
- [5] T. W. Nattkemper, T. Hermann, W. Schubert, and H. Ritter, "Look & listen: Sonification and visualization of multiparameter micrographs," in *25th Annual International Conference of the IEEE Engineering in Medicine and Biology Science (EMBC2003)*, Cancun, Mexico, 2003.
- [6] J. Anderson, "Creating an empirical framework for sonification design," in *11th International Conference on Auditory Display (ICAD'05)*, Limerick, Ireland, 2005, pp. 393-397.
- [7] S. Haykin and Z. Chen, "The cocktail party problem," *Neural Computation*, vol. 17, no. 9, pp. 1875-1902, 2005.
- [8] J. Anderson, "Sonification design for complex work domains: Streams, mappings and attention," 2005.
- [9] A. S. Bregman, *Auditory scene analysis*, The MIT Press, Cambridge, London, England, 1990.
- [10] G. Lorho et al., "Feasibility of multiple non-speech sounds presentation using headphones," in *7th International Conference on Auditory Display (ICAD'01)*, Espoo, Finland, 2001, pp. 32-37.
- [11] H. Hawkins and J. Presson, "Auditory information processing," in *Handbook of perception and human performance*, KR Bott, L Kaufman, and JP Thomas, Eds., vol. II of *Cognitive processes and performance*, pp. 26-64. Wiley, New York, 1986.
- [12] I. Biederman, "Recognition-by-components: A theory of human image understanding," *Psychological Review*, vol. 94, pp. 115-147, 1987.
- [13] V. Best et al., "The effect of auditory spatial layout in a divided attention task," in *11th International Conference on Auditory Display (ICAD'05)*, Limerick, Ireland, 2005, pp. 17-22.
- [14] D. K. McGookin and S. A. Brewster, "Advantages and issues with concurrent audio presentation as part of an auditory display," in *12th International Conference on Auditory Display (ICAD'06)*, UK, 2006.
- [15] B. Shinn-Cunningham and A. Ihlefeld, "Selective and divided attention: Extracting information from simultaneous sound sources," in *10th International Conference on Auditory Display (ICAD'04)*, Sydney, Australia, 2004.
- [16] V. Best et al., "Two-point discrimination in auditory displays," in *9th International Conference on Auditory Display (ICAD'03)*, Boston, MA, USA, 2003.
- [17] B. Arons, "A review of the cocktail party effect," *Journal of the American Voice I/O Society*, vol. 12, pp. 35-50, 1992.
- [18] J. A. Michon and J. L. Jackson, "Attentional effort and cognitive strategies in the processing of temporal information," *Annals of the New York Academy of Sciences*, vol. 423, pp. 298-321, 1984.
- [19] S. H. Han and M. S. Kim, "The role of spatial working memory in visual search efficiency," *Psychonomic Bulletin & Review*, vol. 11, pp. 275-281, 2004.
- [20] G. Woodman and S. Luck, "Visual search is slowed when visuospatial working memory is occupied," *Psychonomic Bulletin & Review*, vol. 11, pp. 269-274, 2004.
- [21] G. Repovs and A. Baddeley, "The multi-component model of working memory: Explorations in experimental cognitive psychology," *Neuroscience*, vol. 139, pp. 5-21, 2006.
- [22] J. C. Webster and P. O. Thompson, "Responding both of two overlapping messages," *Journal of the Acoustic Society of America*, vol. 26, no. 1, pp. 396-402, 1954.
- [23] B. C. J. Moore, *An introduction to the psychology of hearing*, Elsevier Science, USA, 2003.
- [24] T. L. Bonebright, "Perceptual structure of everyday sounds: A multidimensional scaling approach," in *7th International Conference on Auditory Display*, Espoo, Finland, 2001.
- [25] M. Puckette and D. Zicarelli, "Max/MSP," 1990-2008.
- [26] J. M. Jot and O. Warusfel, "A spatial processor for musicians and sound engineers," 1995.
- [27] K. M. Franklin and J. C. Roberts, "Pie chart sonification," in *Information Visualization*, Ebad Banissi, Ed. 2003, pp. 4-9, IEEE Computer Society.

- [28] B. Gardner and K. Martin, "HRTF measurements for a KEMAR dummy-head microphone," May, 1994-1994.
- [29] H. F. Pollard and E. V. Jansson, "A trisstimulus method for the specification of musical timbre," *Acustical*, vol. 55, pp. 162-171, 1982.
- [30] S. McAdams, "Recognition of sound sources and events," in *Thinking in sound: The cognitive psychology of human audition*, S. McAdams and E. Bigand, Eds., pp. 146-198. Oxford University Press, Oxford, 1993.
- [31] J. Krimphoff, S. McAdams, and S. Winsberg, "Characterization of the timbre of complex sounds. II Acoustical analyses and psychophysical quantification," *Journal de Physique*, vol. 4, no. C5, pp. 625-628, 1994.
- [32] J. G. Neuhoff et al., "Pitch change, sonification and musical expertise: Which way is up?," in *8th International Conference on Auditory Display (ICAD'02)*, Kyoto, Japan, 2002.
- [33] B. N. Walker and L. M. Mauney, "Individual differences, cognitive abilities and the interpretation of auditory graphs," in *2004 International Conference on Auditory Display (ICAD'04)*, Sydney, Australia, 2004.
- [34] P. Zahorik et al., "Localization accuracy in 3-D sound displays: The role of visual-feedback training," 2003.
- [35] B. G. Shinn-Cunningham et al., "Perceptual plasticity in spatial auditory displays," *ACM Transactions on Applied Perception (TAP)*, vol. 2, no. 4, pp. 418-425, 2005.
- [36] B. A. Wright and M. B. Fitzgerald, "Sound-discrimination learning and auditory displays," in *9th International Conference on Auditory Display (ICAD'03)*, Boston, MA, USA, 2003.

Hong Jun Song received a B.Eng. in Computer Science and Engineering from Zhengzhou University, China in 1996 and Master's degree in Internet Technology from the University of Wollongong, Australia in 2003. She is currently pursuing a Ph.D. Degree at the Key Centre of Design Computing and Cognition at the University of Sydney, Australia.

Her interests include the development of sonification (delivery of information by mapping data onto auditory dimensions) and the exploration of effective information sonification design strategies to facilitate intuitive comprehension.

Dr. Kirsty Beilharz is a Senior Lecturer in Digital Media at the Key Centre of Design Computing and Cognition, Faculty of Architecture, Design and Planning at the University of Sydney. Her research includes gestural interaction with sonification, wearable sonification and context-aware informative display. Her teaching includes Sound Design and Sonification and an Interactive Sound Design Studio.