Spatialization and Timbre for Effective Auditory Graphing

HONG JUN SONG and KIRSTY BEILHARZ
Key Centre of Design Computing and Cognition
Department of Architecture, Design and Planning
University of Sydney
AUSTRALIA

Abstract: Spatialization (spatial separation) and timbre are the two salient auditory attributes facilitating a better comprehension of sonification (non-speech auditory display).While directional cues provide obvious interpretable structure for auditory display, timbre provides semantic expression by highlighting identity of auditory stream. This paper aims to explore (1) the impact of spatial separation for a divided attention task and (2) the efficiency of timbre to assist contour identification. The observation was applied to auditory graphing, which can be used to represent a wide variety of numerical data sets. The results help provide empirical evidence, contribute to a further investigation of spatialization and timbre employed within an auditory display context of a real-world scenario (e.g. sonification of digitalized social data).

Key–Words: Spatialization, Directional separation, Timbre, Auditory graph, Divided attention, HRTF, Binaural

1 Introduction

Auditory graphs or sonified graphs are broadly defined as a means to display quantitative data (e.g. scientific data or real-time financial data) [1]. The numerical data values are mapped onto pitches on the y-axis of auditory graphs and the x-axis corresponds to the movement of the pitch over time. Visualization of dense and complex data sets normally has dense overlapped graphical lines or clusters. The use of non-speech sound (sonification) to display these quantitative data has the potential to relieve visual overload as an alternative or supplement to visual display. However, the principles for auditory graphs are not as well documented as visual graphs and designing a good auditory graph is associated with effective methods of displaying data. When working with multi-variation quantitative data, difficulties arise because of the interaction between parameters/dimensions (pitch, loudness and rhythm etc.), interference of audio streams (e.g. masking) and multi-tasks (e.g. divided attention). This paper aims to examine the design choice for perceptually promoting auditory grouping and to contribute to the body of empirical evidence.

Representing multi-dimensional data through auditory display is challenging. Depending on the numbers of auditory streams being recognized, the tasks are compartmentalized into divided and selective attention. In divided attention both of the concurrent signals need to be identified and in selective attention only one important audio stream needs to be extracted from competing auditory messages [2, 3, 4]. The principles about 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 [5]. In multi-stream situations, it is easily to cause confusion and blend of auditory streams. Thus mapping strategies have to meet both the requirements of "informativeness" (of quantitative data) and "simplicity/clarity" (for the listener). Therefore, the primary investigation is how sound dimensions 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.

2 Directional cues for stream segregation

Spatial separation has a better "force and semantic structure" than pitch to reduce "problems of peripheral sensory masking" [6] and maintain attentions to 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 [7]. 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 about "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 (heard as pitch and loudness respectively). Secondly, interaural time/level differences reduce masking and interference between streams [8] and improve the signal-to-masker ratio and the audibility of the target that allow listeners to direct their attention to the target.

While spatial location is not frequently discussed like pitch among the available auditory parameters, we advocate that it is an effective dimension for auditory display [9]. In the study of Lorho et al.[9], 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 [10]. Furthermore, research in speech field consistently shows the benefits of spatial separation for splitting competing verbal messages displayed by headphones or loudspeakers [8, 11, 12]. Cocktail party effect 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 [13]. While intuition indicates that for non-speech sound localization is able to enhance auditory stream segregation, other studies show that the effectiveness of spatial separation is restricted for categorical data [13] and for divided attention situations, the temporal information about individual strands of two auditory graphs will be destroyed due to individual's spatial cognitive ability [14].

3 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" [15]. Concerning the features of timbre, many studies illustrate the important role of timbre for stream segregation, such as auditory scene analysis (ASA) [5]. Furthermore, timbre plays a crucial role for an effective auditory display. It is essential to determine the acoustic variables 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 them, it facilities the creation of desired sounds concerning timbral quality and comprehensibility. For example, the study done by Bonebright [16] 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 the spatialization. The analysis consists of correlation between sound quality and graphs, acoustic measurements, and inter-correlations between these two sets of variables together.

4 Experiment design

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

This experiment investigated the participants' ability to spatially identify the contours of two concurrent auditory graphs when their 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 measuring 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. This experiment also investigated the timbral quality of fourteen synthesis sounds. 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.

4.1 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, 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.

4.2 Stimuli

The auditory stimuli were created in Max/MSP [17]. Max/MSP is a real time graphical programming environment. The parameters such as playing tempo can be easily altered 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 [18] supports the HRTF function for binaural synthesis. 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 [18].

The experiment used headphones for binaural reproduction with generic HRTFs. No head-tracking was involved. While this simple binaural presentation technology has the two problems of cone-of-confusion errors (such as front-back confusion)

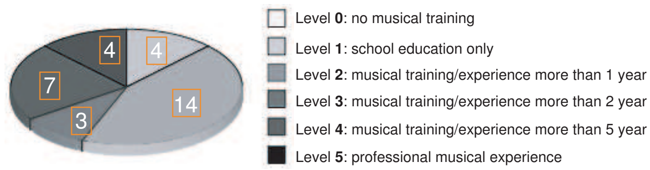


Figure 1: *Pie chart of musical background (MBG). Six levels were rated from beginner to expert. Numbers of people at each level are marked in corresponding pie segments. No participants had professional musical training, so the pie chart does not include this segment.*

Table 1: *Stream configuration.*

	Within a pair		Between pairs	Between modes
Parameters	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	-	changing (0 or +/-60)
virtual distance	constant	constant	-	-

and head-locking of the sound-field (the sound-field moves with the listener's head), the spatial separation used for the stimuli 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.

The configuration of auditory graphs within and between streams/pairs/modes is illustrated in Table 1. 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.

4.2.1 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 explore the dependence of co-variation and coincidence between graphical lines and percentage of errors. The pairs of lines were marked as "-1", "0" and "+1", corresponding to negative, null and positive correlation

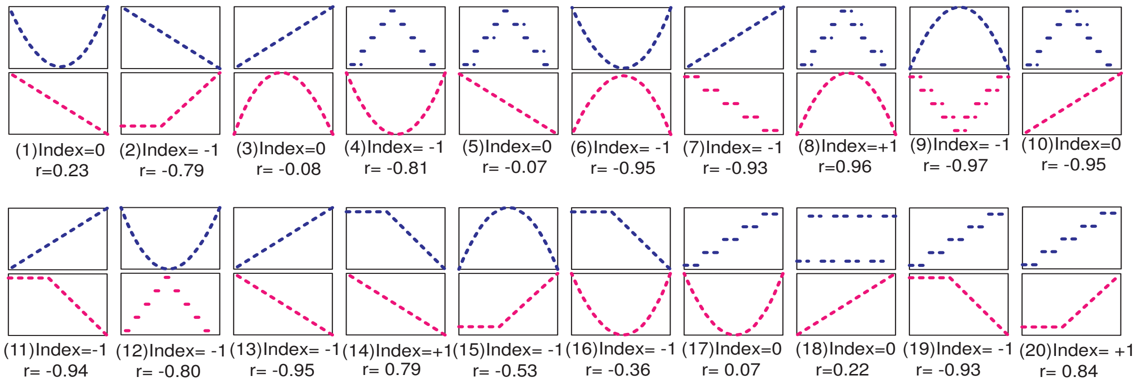


Figure 2: *Twenty pairs of graphs for sonification numbered from (1) to (20). Values of each graph were mapped onto pitches and each graph represents one audio stream. The index shows the correlation between two auditory graphs where "-1": negatively correlated, "+1": positively correlated and "0": no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

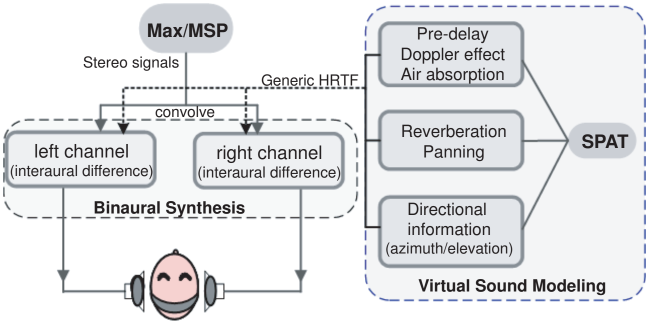


Figure 3: *System Overview. The two core units consist of virtual sound modeling and binaural synthesis, which render the virtual realism and realistic sound. Pre-delay (Doppler Effect, air absorption and reverberation) and pre-equalization are accomplished by SPAT. Generic HRTF are used to enhance binaural presentation.*

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 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 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.

4.2.2 Reproduction of virtual spatial location

Best et al. found that interference still occurred up to 60 separation of verbal messages [12]. According to the previous findings [15, 19] for concurrent minimum audible angle (CMAA), our case used 120 separation two concurrent sound sources. 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 to the left of the middle line and -60 to the right of the middle line (all at 0 elevation and 1m distance) (Figure 3). The two different display modes are shown in Figure 2. The HRTFs codes were collected by MIT's Media Laboratory [20]. 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.

4.2.3 Timbre generation with FM/AM synthesis

Fourteen timbres were generated applying for different pairs of auditory graphs (Table 2). 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 envelope was shaped by amplitude modulation (AM) with alteration of graphical breakpoints in Max/MSP. The linear envelope was made of several straight lines with various relative durations.

Feature extraction is based on spectral centroid $f(c)$. There are other techniques such as tristimulus (which describes timbre equivalent to colour attributes in vision) [21] and spectral irregularity/smoothness (in which the average of current, next and previous amplitude is compared with current amplitude) [22, 23]. The spectral centroid corresponds to brightness/sharpness of a sound which is perceived brighter with more high frequency components.

Table 2: *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)	

4.3 Task and procedure

The tasks were to listen to the graph sonification containing a pair of simultaneous audio streams (which were played twice with a 7.5s pause between individual) and match the contour of both streams to 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. The first five listening examples were regarded as training and not used in comparison. For each trial, the auditory stimuli consisted of a pair of contours (10 s) and a repetition. 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.

5 Results

The overall performance showed that spatial separation does improve listeners' ability to attend to two competing pitch contours when using binaural representation. The effect of spatializsation was quantified by comparing the difference of performance when the pair of audio streams were spatially separated or co-located. 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 for co-located. The performance of 32 participants with two display modes is summarized in Figure 4.

5.1 Discriminability of spatial separation

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 3.

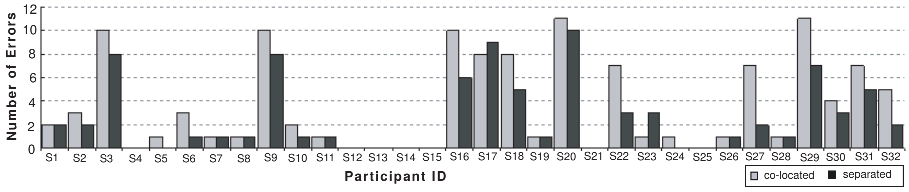


Figure 4: *Summary of occurred errors in co-located and separated modes. Subjects with ID S4, S12, S13, S14, S15, S21 and S25 had made no errors in both of the display modes; S5 and S24 only have errors occurred in co-located mode.*

Table 3: *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
[2] separated sound source) 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. Therefore, their performance [t(31)=3.968, p=0.005<0.01] illustrated 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 when they were 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.

5.2 Influence of graph correlation

Figure 5 plots the results for divided attention tasks as a function of graphical correlation index (corresponding to Table 1). (a) Errors dropped when two signals were spatially separated (dashed line).

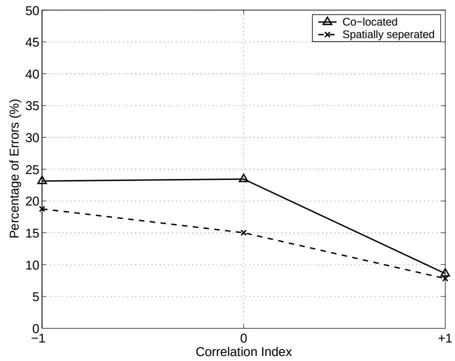


Figure 5: *The results are shown for co-located and separated sound source location (solid and dashed lines respectively), which plots the mean errors in recognizing the pairs of auditory graphs with graphical correlation index. It demonstrates that fewer errors occurred when sound sources of the two 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).*

The deviation is more obvious when they are not correlated ("0") than negative 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 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 reported that the auditory graph pairs combined with steps (e.g. graph (4) in Figure 2) were more easily identified. Since the different performance were found between pairs with steps (26.1% errors) and without (36.8% errors), steps to represent data are considered to be employed in future sonification applications.

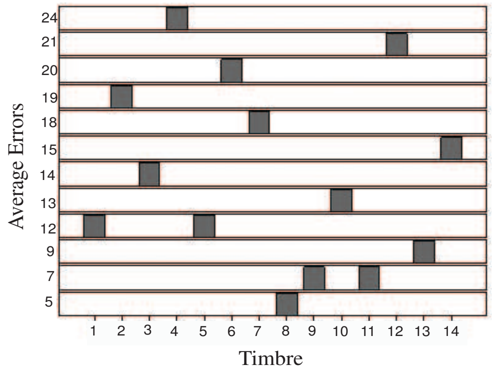


Figure 6: *Average errors of 14 timbres in both spatial display modes. Timbre D, F and L produce the highest number of errors, while timbre H, I and K have the fewest errors.*

5.3 Timbral quality ranking

This paper considers timbre as a dependent parameter and the analysis of its impact was confounded with the pitch contours. By comparison of total errors occurred in both of the display modes, the ranking of average errors for each timbre is shown in Figure 6. 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. 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 (shown in Figure 7). Additionally, further comparison (combining graph correlation) between bottom three shows that H's correlation is "+1" (fewest errors) and the two others' correlation index is "-1". It indicates that positively correlated auditory graphs tend to be easier identified than negatively correlated pairs and auditory graphs with relatively long attack tone color tend to be easier identified than those with relatively short attack tone color. In the questionnaire responses some timbres were reported being perceived as more distinctive than others.

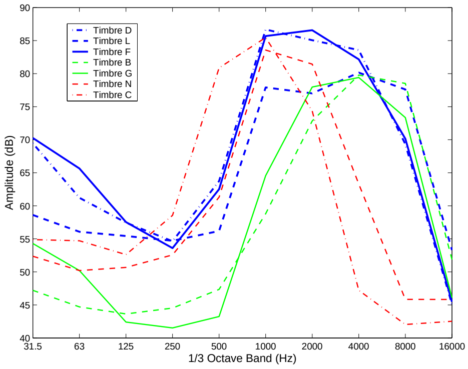
6 Discussion

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 might be 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. 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.

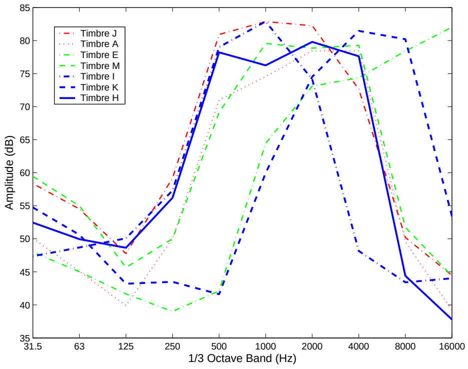
Musical experience was considered as one of the human factors (such as age and gender) related to cognitive ability in many previous studies. In Neuahoff et al's case [24], musical experts and musical novices responded differently when the tasks was pitch magnitude estimation. Prior musical knowledge and expertise contributed to reaction time and the accuracy of the interpretation. 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 the 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 [25]. 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. Participants pointed out that they became familiar with the display of the stimuli and felt comfortable after first a few trials, so the first five trials were excluded from analysis and regarded as training.

One-way ANOVA has found no significance among six levels [$F(4,27)=0.352>0.05$]. 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 influence the pattern recognition. These results are consistent with the findings of Walker and Mauney [25].

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 [$F(1,30)=0.090>0.05$].The Spearman correlation test showed that it is likely



(a)



(b)

Figure 7: *Spectral envelope of 1/3 octave band analysis. Graph (a) includes top 7 timbres in ranking and Graph (b) includes bottom 7 timbres. The spectrum of D, F and L are almost overlapped which are at the top of performance ranking graph (Figure 6).*

that people, who cannot do well in spatially separated display, also cannot perform well in co-located display. The results clarify the effect of directional cues facilitating concurrent auditory graph identification. However, the study of timbre is coarse while listenable (over time) and effective sound is an important consideration in designing sonification for uses in practical situations. The decision of timbre in this study is based on a rough selection from a small set of FM/AM synthesised tone colors. Out next work will focus on the evaluation of synthesised timbres which aims to find out applicable synthesised timbres from a listener's perspective. 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.

7 Future works

The selection of used timbres in the current study was based on a small set of synthesised timbres with primary consideration of dissimilarity. Since sound quality is important in designing sonification for uses in practical situations, emphasis is placed on evaluating the timbre of the synthesised sounds from a listener's perspective, by integrating perceptual characteristics of generated tone colors 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 [26, 27, 28]. 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] 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.

[2] J. Anderson, "Creating an empirical framework for sonification design," in *11th International Conference on Auditory Display (ICAD'05)*, Limerick, Ireland, 2005, pp. 393–397.

[3] S. Haykin and Z. Chen, "The cocktail party problem," *Neural Computation*, vol. 17, no. 9, pp. 1875–1902, 2005.

[4] J. Anderson, "Sonification design for complex work domains: Streams, mappings and attention," 2005.

[5] A. S. Bregman, *Auditory scene analysis*, The MIT Press, Cambridge, London, England, 1990.

[6] 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.

[7] I. Biederman, "Recognition-by-components: A theory of human image understanding," *Psychological Review*, vol. 94, pp. 115–147, 1987.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] V. Best et al., "Two-point discrimination in auditory displays," in *9th International Conference on Auditory Display (ICAD'03)*, Boston, MA, USA, 2003.

[13] B. Arons, "A review of the cocktail party effect," *Journal of the American Voice I/O Society*, vol. 12, pp. 35–50, 1992.

[14] 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.

[15] B. C. J. Moore, *An introduction to the psychology of hearing*, Elsevier Science, USA, 2003.

[16] T. L. Bonebright, "Perceptual structure of everyday sounds: A multidimensional scaling approach," in *the 7th International Conference on Auditory Display*, Espoo, Finland, 2001.

[17] M. Puckett and D. Zicarelli, "Max/msp," 1990-2007.

[18] J. M. Jot and O. Warusfel, "A spatial processor for musicians and sound engineers," 1995.

[19] K. M. Franklin and J. C. Roberts, "Pie chart sonification," in *Information Visualization*, Ebad Banissi, Ed. 2003, pp. 4–9, IEEE Computer Society.

[20] B. Gardner and K. Martin, "Hrtf measurements for a kemar dummy-head microphone," May, 1994 1994.

[21] H. F. Pollard and E. V. Jansson, "A tristimulus method for the specification of musical timbre," *Acustical*, vol. 55, pp. 162–171, 1982.

[22] 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.

[23] J. Krimphoff et al., "Characterisation du timbre des sons complexes. ii: Analyses acoustiques et quantification psychophysique.," *Journal de Physique IV*, vol. 4, no. C5, pp. 625–628, 1994.

[24] 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.

[25] 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.

[26] P. Zahorik et al., "Localization accuracy in 3-d sound displays: The role of visual-feedback training," 2003.

[27] 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.

[28] 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.