

Development of Sonification Strategies Facilitating Auditory Display

Hong Jun Song **Kirsty Beilharz**
Key Centre of Design Computing and Cognition
Faculty of Architecture, Design and Planning
University of Sydney
hson5790, kirsty@arch.usyd.edu.au

ABSTRACT

Information sonification is an emerging field of pervasive computing and is applicable in many wearable artificial sensory systems. This paper presents some strategies that can be employed in auditory interface design where using sound as representation gives insight into the improvement of comprehension and recognition of information sonification in an auditory graphing system. The design strategies are based on listening tests gathering contour identification responses and listening experiences from participants. The analysis of these responses from identification testing and surveys then suggests the method of the usage of auditory dimensions in auditory display. The empirical evidence benefits the development of sonification design strategies for auditory interfaces that can be extended to a wide range of interactive wearable applications. Our discussion will include potential applications of this knowledge in the fields of ambient information display and wearable auditory or audio-visual displays, i.e. ways in which timbre and spatialization can enhance the identification and discriminability of sonified data streams.

Keywords

Information sonification, spatialization, timbre, quantitative measurement, qualitative measurement, aesthetic display

INTRODUCTION

In everyday life people communicate and collect information by using all their physical senses, including vision, hearing and touch. Sound is regarded as an informative medium to support an information-seeking task or activity. For example, many devices such as doorbells, ovens and telephones use sounds (auditory cues and auditory alerts) to communicate events, environment, and gain awareness. Sounds are a natural consequence of the physical environment and although being seldom aware of it, people are always listening and using sounds in many ways [1]. Also due to its intuitive connection with symbolic meanings, sound has been extensively used as a means for Human Computer Interaction (HCI), in which it can provide a rapid comprehension of the processing status, alarm or warnings in the user interface. In novel HCI paradigms such as ubiquitous, pervasive and wearable computing, interactive sonification provides alternatives to

visual display, intuitive responses and augmented channels of communication [2, 3]. Apart from its accessibility for visual impaired users [4], sonification can also be used for navigation cues [5], representing complex scientific data such as neural spike propagation data [3] (the reference gives an example where a group of players can collaborate with each other and their actions transferred through a percussive controller were responded to composed pieces of music), forecasting real-time financial data [6, 7], and other responsive real-time data, such as capturing dynamic non-visual weather data [8].

Taking account of humans' pervasive experience with sound and intuitive response to sound, representing information in an auditory display is intended to aid and enhance comprehension of the data represented, and to meet the needs of extending computational data processing with the fast growth of digitalization in our society. Some multi-dimensional sets of data benefit from auditory representation (sonification) because the time-based, linear representation of sound allows us to easily recognize trends, patterns, information clusters and recurrences of behaviour. This leads to easier assimilation of the displayed information, especially in visually overloaded workplaces.

However, because of the instantaneous features of sound, auditory interface design appears to be challenging. For sonification to be of optimal benefit, the rapid cognition, comprehension and assimilation of auditory information in a non-demanding, intuitive representation are essential to its efficacy.

AIM AND MOTIVATION

This study is motivated by a desire to systematically examine how spatial separation and timbre assist on the ability of listeners to understand sonification, taking the advantage of pervasive listening to sound. It aims to explore sonification design strategies that facilitate efficient information representation and that have great applicable

potentials in various wearable synthetic systems.

An important issue in auditory interface design is to convey the intent of the designer and present an easily-understood auditory display for users. In information sonification, various types of information need to be displayed clearly and unambiguously, and some obstacles have to be addressed for an efficient representation. In order to achieve the goal, sonification research involves multidisciplinary investigation in the fields of acoustics, psychoacoustics, psychology, cognition and human computer interaction. It is a significant challenge to determine how to design a “comprehensible” auditory display and to map information onto adaptive sound variables, because of the lack of applied theories of mapping generation for a successful sonification. The gap can be bridged by comprehensively investigating the essential components in sonification design.

Mapping is a core process in auditory interface design to establish linkages between the data domain and the sound domain, e.g., connecting data quantity, activity or numerical values to auditory attributes such as pitch, rhythm and location. This computational process determines to a high degree how successfully a listener is able retrieve the information defined in the mapping process. In other words, the mapping of data onto sound requires a deep understanding of the way of information retrieval by a listener – the recognition of information embedded in sound – and must take into account the user’s capability while carrying out tasks in the real environment.

EXPERIMENT I – USAGE OF SPATIAL SEPARATION TO IMPROVE IDENTIFICATION AND RECOGNITION

When two sound sources are oriented from different directions, a listener needs to keep tracking and integrating auditory messages from two spatially separated locations. Spatial rehearsal allows tracking the orientation locations in the working memory and spatial reasoning is augmented due to the identification of the masking and interference occurring between two auditory streams. It is possible to degrade the performance if an interactive interface designer uses this strategy (analogous to the findings in visual display [9, 10]). In other words, cognitive overload and complexity may hinder performance.

While intuition may indicate that spatial separation can be used in the mapping scheme of auditory interface design [11], due to its feature of having less interaction with other parameters such as pitch and rhythm and interaural time/level differences [12, 13], there is little empirical information in literature supporting and investigating the effects of non-speech spatialization.

Experiment I especially examined the possibility of spatial separation used as mapping strategy to assist listeners during the deciphering information embedded in sound. Our concern is particularly to develop an understanding of the potential application of spatialization in sonification

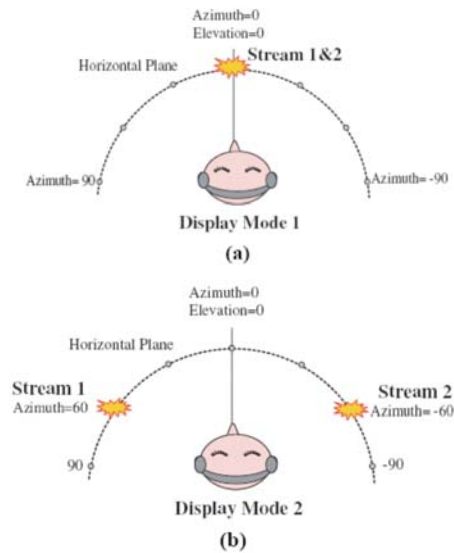


Fig. 2. Spatial configuration in experiment I. Mode 1: both streams are from the middle line (azimuth= 0°and elevation=0°); Mode 2: they are spatially separated, 60 degrees away from the middle line symmetrically.

using commercially available binaural reproduction. The results from this experiment confirmed the benefits of spatial separation in information sonification with headphone representation.

Method

Forty stimuli were included for the 32 volunteer participants and their task was to identify the contours of pairs of various auditory graphs. The pairs of simultaneous auditory graphs were configured both in a single location and in two separated virtual locations (see Figure 1). The spatial rendering was simply in terms of image direction, without any attempt to vary the virtual distance or other aspects of auditory space. The combination of 20 stimuli in display mode 1 and 20 stimuli in mode 2 was played in a random sequence to the participants and the subjects’ task was to pick up the correct pair of contours from 12 given options.

The experiment used headphones for binaural reproduction with generic Head-related Transfer Functions (HRTFs) and no head-tracking was involved. We recognized that the accuracy of head-tracking, individualized HRTFs or loudspeaker reproduction is greater, but these laboratory conditions do not simulate the workplace, such as the stock trading data analysis context and are not applicable for pervasive computing systems.

The auditory interface (shown in Figure 2), creating stimuli for the listeners, are designed in Max/MSP [14], which is a real-time programming environment. Parameters such as the energy envelope can be easily altered by editing the graphic icons.

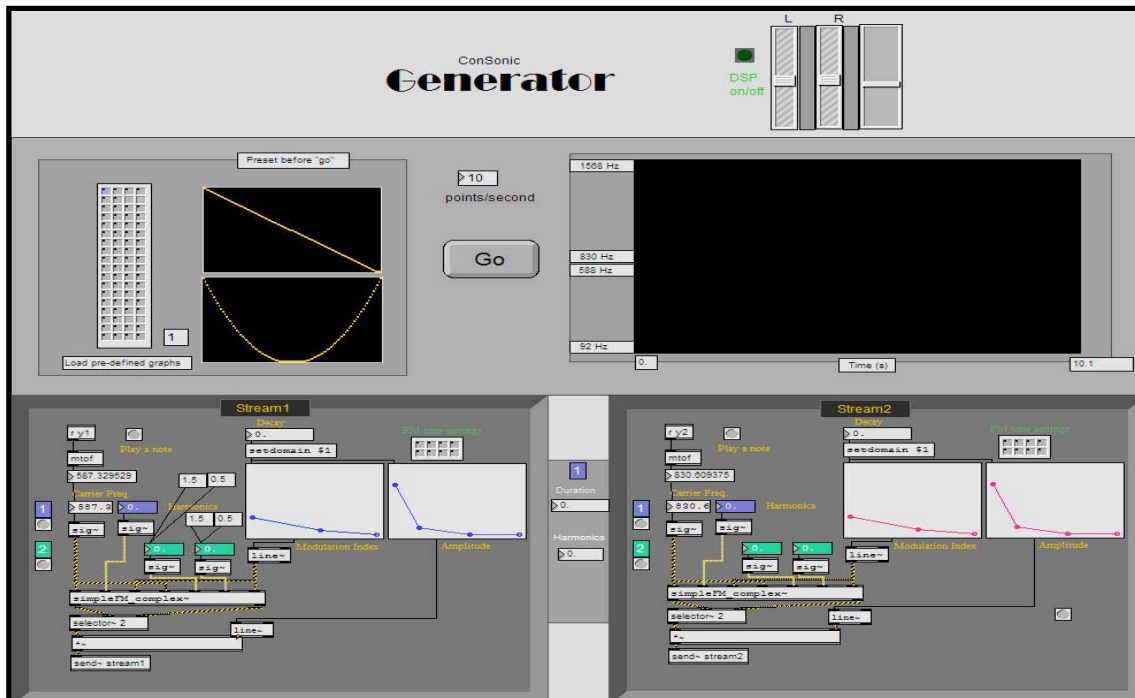


Fig. 3. Screenshot of the synthesizer to create auditory graphing stimuli. The up-left module was a group of preset graphical lines and the shape can be displayed in the up-right screen visually. The two modules at the bottom are able to generate different timbres for pairs of timbres.

Results

When the participants identified the pairs of contours, a significant difference ($P = 0.005 < 0.01$) was found between the two display modes. When the two auditory streams were spatially separated, the errors that occurred were fewer than when the same pair was co-located.

The result indicates that spatial separation can be used as a mapping strategy in auditory interface design where it is able to improve the comprehensibility, when a listener has to pay attention to two simultaneous auditory events. The other outcomes are summarized as:

- The human factor of musical background did not influence a listener's performance where the listening task was about sketching relative differences (i.e., pitch contour) rather than judging absolute differences (i.e., magnitude estimation – degree of pitch difference).
- Timbre has the potential to enhance spatialization in sonification. Although in this experiment, pairs of auditory streams used the same timbres, differences were found between pairs with different timbres. The analysis of acoustic attributes showed that the timbres with a long attack performed better than those with a short attack.

EXPERIMENT II – USE OF INDIVIDUAL TIMBRE TO ENHANCE SPATIALIZATION AND AESTHETICS

The usage of timbre in sonification is not an easy task because timbre is not defined by a single value but by a

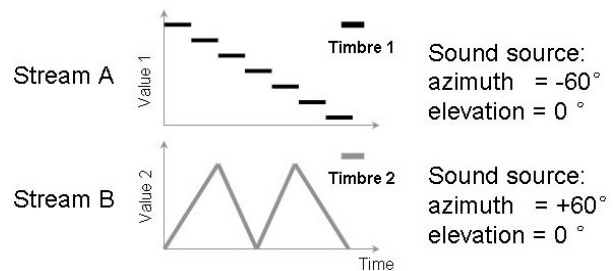


Fig. 5. Sample stimulus. The stimulus consisted of two concurrent auditory streams. Numerical values of graphical lines were mapped onto pitches on the y-axis and the movement of pitch over time was along the x-axis. Each stream had its own timbre and sound location (60 degrees away from the middle line). Therefore, stream segregation benefits from both timbral quality and spatial separation.

group of dynamic and static sound properties [15]. In music and speech, timbre is the quality that distinguishes different types of sounds of musical instruments or voices. In sonification, it is regarded as a remarkable dimension [16] that allows characterization, distinction and comparison of sounds. Additionally, timbre is an important variable when presenting an aesthetic auditory display.

Previous studies [17, 18] on timbre tried to characterize timbre in a perceptual space. These investigations contribute to a better representation of timbre by restricting the number of dimensions to three perceptually related dimensions. This assists in the determination of auditory variables to generate compatible timbres in this experiment.

Different from timbre space studies, experiment II augmented a measurement of timbre for a cognition task – in order to understand whether timbre assists recognition of contours.

Experiment II aims to explore key factors influencing the use of timbre when designing an auditory interface. The evaluation was performed from the angles of “informativeness”, “aesthetics” and “clarity”. The selection of “distinct” pairs of timbres were based on the features that are not only able to differentiate concurrent auditory streams but also able to aid the identification of contours. The decision about the mapping strategy, concerning timbre, was derived not only from the survey about the participants’ listening experiences but also from identification performance.

Method

The timbres used were created with Fourier-based additive synthesis. Each auditory stream has its own individual timbre (see Figure 3). The graphical values were mapped onto fundamental frequencies and the output was a sequence of digital signals, each of which was a summation of the fundamental and the harmonics/partials. The generation of timbre was controlled by altering the attack portion, ratio and number of harmonics/partials, and amplitude envelope. Additive synthesis also allows an evolutionary change in timbre with the change in frequencies reflecting numerical data values.

Thirty-six volunteer participants, who ranked themselves in the post-experiment survey as music novices, in terms of their musical background, were included in experiment II. They had only one chance in the identification task to listen to each stimulus in order to avoid an incorrect listening strategy – listening to one stream and then the other when replaying. But when accomplishing a semantic differential scaling task, the stimuli can be played as many times as the listeners required.

A semantic differential technique was used to compare and identify the degree of sensation of distinctiveness of pairs of synthesized timbres and the sensation of the sound quality of selected outstanding auditory streams. These bipolar scales define the strength level of each feeling. The sound quality scales were rated from five perspectives of subjective listening experiences: relaxation (focused/diffuse), aesthetics (smooth/rough), clarity (obvious/subtle and soft/hard) and psychoacoustics (sharp/dull). Using a set of rating scales derived from the semantic differential method, the subjects judged and described each of five different feelings in terms of their own emotional impressions.

Acoustic measurement was also involved. The analysis was predominantly in the spectral domain and was based on feature extraction from six perspectives: number of harmonics, average intensity, peak frequency, ceiling amplitude, spectrum width and spectral centroid.

Results

Combining the performance and the listeners’ ranking of the distinctiveness between pairs of timbres, a significant correlation was shown ($p = 0.014 < 0.05$). This indicates that in auditory interface design, auditory streams could have their individual timbres because distinct pairs of timbres can provide assistance to both stream segregation and information recognition.

Semantic differential scaling illustrates underlying listeners’ underlying emotional meanings to sound. The verbal descriptors portrayed the main characteristics of the striking/noticeable auditory stream of a pair. These selected streams have individual prominent timbral features that attract the listeners’ attention and perceptually separate them from simultaneous streams. Therefore, the descriptors of “obvious” and “focused” are most significant and consistent. The correlation between smooth/rough and soft/hard indicates that aesthetics and relaxation characteristics are related and “smooth” sounds have less perceptual tensivity.

Acoustic measurements illustrate that performance is affected by the spectrum width and spectrum centroid of the pairs of timbres.

Integrating the performance, semantic differential scaling and acoustic measurements, the outcomes are summarized below:

- It is shown that distinctiveness of timbres is a feature that can enhance the performance of a spatialized information sonification in the contexts of concurrency, comparison and recognition.
- Distinct timbre pairs can help not only stream segregation (perception only) but also contour identification (cognition).
- Sharpness of sound makes an auditory stream perceptually prominent, which helps stream segregation but does not help contour identification. Therefore, for a better sonification display, the concurrent auditory streams should be differentiated in timbre but the differentiation should be achieved using the property of sharpness.
- Acoustic measurement shows that the differentiation can be achieved by enlarging the gap of spectrum width and spectral centroid between two auditory streams, but not by increasing ceiling amplitude and peak frequency.

DISCUSSION AND CONCLUSION

The results of the two experiments, which gathered the responses and listeners’ opinions in a real environment, contribute to the development of strategies for ubiquitous interactive auditory interface design. The study presented can be applied into a wide range of interactive interfaces and could be incorporated with visual and haptic devices to develop creative information display systems.

These strategies can be especially applied in a complex auditory display including concurrent auditory display where a listener has to pay attention to and recognize the pattern of both auditory streams. In information sonification, although single stream representation is relatively easier to follow and recognize compared with multiple streams, it has limitations which restrict bandwidth and the competence of communicating with information. For two time-based data sets, for a single stream display either by a compound of two data sets (i.e., differentiating in timbre), or individually in succession, the time cue is easily lost and comparison (finding differences or relationships between two streams) becomes very difficult. A display containing concurrent auditory streams unties the restriction of single stream display, allows consecutive processing of multiple equally important auditory streams and carries more information. In order to overcome the difficulties of concurrent display such as interference and masking when two streams are harmonically related [19], the differentiation strategies found are able to distinguish streams in order for people to have an adequate understanding of the sonification.

In specific relation to application in wearable expression, e-fashion and mobile body-bound contexts, spatialization of audio faces some restrictions. Small wearable displays might be limited to a single speaker or loudspeaker sound production systems that offer little spatial separation between channels or to binaural (headphone) display. In this case, our findings about timbral spectra and differentiating streams for identification and separation will be most useful. As time-based audio display is often continuous, a cue-based or alert system might better suit a wearable display. Spatialisation adds a dimensions for sonification that, when worn, might imbue the wearable with a sense of directionality. However, in an architectural (interactive architectural or responsive architectural) or sensate context, such as for ambient information display, bi-modal and interactive sonifications or installations in public spaces, **both** spatialization and timbral differentiation can enhance the understanding of auditory information representation. Panning, for example, is easily achievable in a display with well-separated speakers. In ambient contexts where the listener is likely to encounter interference of ambient noise and complexity created by environmental, conversational and other competing sounds, strategies that enhance stream identification and comprehension of concurrent data encoding contribute to usability and sustained listening for the user. Further, most existing sonification toolkits and time-based displays emphasize the dimensions of pitch and periodicity for representing data attributes and we believe the findings in this paper support both the validity and advantage of using other (additional) sonic parameters to express information. Subtleties such as timbral nuance and a rich array of location sources in the spatial image arguably contribute to a deeper emotive and persuasive quality of sonification. In particular, we aim to give greater importance to the

aesthetic aspects of auditory display for social and community contexts, so that people enjoy wearing or inhabiting the display. In regard to the potentially expressive nature and aesthetics of wearable displays, the additional subtlety and dimensionality afforded by using an array of timbral (tone quality) variations is both limitless and more textural or sensitive than only using pitch and periodicity. We think a parallel might even be drawn between the textural nuances of timbre and the materiality of textiles, relating qualities of surface, smoothness, roughness, suppleness, distortion of sound quality and fabric character.

Our current and continuing work extends these findings to time-based data and emphasizing qualitative measurement, which is able to uncover the interaction between listeners and auditory display and to reflect on the “process” through which a listener experiences decoding sounds.

REFERENCES

1. Barrass, S. (1997). Auditory Information Design, *PhD degree dissertation*, Australian National University
2. Rath, M. and Rocchesso, D. Continuous Sonic Feedback from a Rolling Ball, *IEEE Multimedia Special Issue on Interactive Sonification*.
3. Weinberg, G. and Thatcher, T. (2006). Interactive Sonification: Aesthetics, Functionality and Performance, *Leonardo Music Journal*, **16**.
4. vOICe 2004 Draw Your Own Sound: Image to Sound Mapping <http://www.visualprosthesis.com/javoice.htm> Aug-15-04.
5. Lokki, T., Grohn, M., Savioja, L. and Takala, T. (2000). "A Case Study of Auditory Navigation in Virtual Acoustic Environments", *9th International Conference on Auditory Display (ICAD'03)*, Boston, MA, USA.
6. Ciardi, F. C. (2004). "Smax a Multimodal Toolkit for Stock Market Data Sonification", *10th International Conference on Auditory Display (ICAD'04)*, Sydney, Australia.
7. Ben-Tal, O. and Berger, J. (2004). Creative Aspects of Sonification, *Leonardo, ISAST*, **37** (3): 229-232.
8. Paine, G. (2004). "Reeds: A Responsive Sound Installation", *International Conference of Auditory Display (ICAD)*, Sydney, Australia.
9. Han, S. H. and Kim, M. S. (2004). The Role of Spatial Working Memory in Visual Search Efficiency, *Psychonomic Bulletin & Review*, **11**: 275-281.
10. Repovs, G. and Baddeley, A. (2006). The Multi-Component Model of Working Memory: Explorations in Experimental Cognitive Psychology, *Neuroscience*, **139**: 5-21.
11. Best, V., van Schaik, A., Jin, C. and Carlile, S. (2005). Auditory Spatial Perception with Sources Overlapping in Frequency and Time, *Acta Acustica United with Acustica*, **91**: 421-428.
12. Best, V., Ihlefeld, A. and Shinn-Cunningham, B. (2005). "The Effect of Auditory Spatial Layout in a

- Divided Attention Task", *11th International Conference on Auditory Display (ICAD'05)*, Limerick, Ireland, pp. 17-22.
13. Shinn-Cunningham, B. G. (2005). "Influences of Spatial Cues on Grouping and Understanding Sound", *Forum Acusticum 2005 Budapest*.
 14. Puckette, M. and Zicarelli, D. (1990-2008). Max/Msp, Cycling74/IRCAM, URL: <http://www.cycling74.com/>
 15. Loureiro, M. A., Paula, H. B. d. and Yehia, H. C. (2004). "Timbre Classification of a Single Musical Instrument", *the 5th International Conference on Music Information Retrieval (ISMIR 2004)*, Barcelona, Spain.
 16. Bregman, A. S. (1990). *Auditory Scene Analysis*, The MIT Press, Cambridge, MA, USA.
 17. Terasawa, H., Slaney, M. and Berger, J. (2005). "Perceptual Distance in Timbre Space", *11th International Conference on Auditory Display (ICAD'05)*.
 18. Grey, J. M. (1977). Multidimensional Perceptual Scaling of Musical Timbres, *the Journal of the Acoustical Society of America*, **61** (5): 1270-1277.
 19. Best, V., Gallun, F. J., Carlile, S. and Shinn-Cunningham, B. G. (2006). Binaural Interference and Auditory Grouping, *Journal of the Acoustical Society of America*, **121** (2): 1070-1076.