

The Role of Auditory Attention and Auditory Perception in the Design of Real-Time Sonification of Anaesthesia Variables

Janet Anderson, Penelope Sanderson and Michael Norris
ARC Key Centre for Human Factors and Applied Cognitive Psychology
University of Queensland, St. Lucia, Brisbane, Australia
Email: janderson@humanfactors.uq.edu.au

Keywords: sonification, auditory attention, auditory perception, auditory display

Abstract

Sonification—the representation of data relations in sound relations—is attracting increasing attention within the human factors community as a way of providing human operators of real-time processes with continuous information about the state of a system. This is particularly needed in anaesthesia, where the anaesthetist must divide attention across a wide variety of tasks. Efforts to design effective sonifications of the physiological state of anesthetized patients, however, have not emerged from basic scientific studies of auditory attention. Literature reviews indicate that the kind of basic scientific research on auditory attention that would support the design of sonifications does not exist. The work described herein is the first part of a research program in which we build a foundation for our design of anaesthesia sonification in a series of basic studies of auditory attention. These studies may contribute information that will support sonification design beyond the anaesthesia application.

1 Introduction

Anaesthetists operate in a domain that is both dynamic and complex. Workload can vary from low to extremely high, especially if there is a high-risk situation to deal with (Gaba & Howard, 1995). The anaesthetist must monitor many physiological variables to ensure that the patient's major organs and physiological systems are not damaged during anaesthesia. These variables are constantly changing, even when they are within the normal range. Complexity arises because there are multiple possible causes of problems and multiple possible interventions. Continuous auditory presentation of data has a number of advantages over visual presentation in providing support for the anaesthetist who is dealing with this complexity. When the information load is high, the operator may not be physically able to monitor all visual displays at a high enough frequency to detect problems. Auditory displays access an additional information channel and allow attentional resources to be distributed across modalities. Communicating information in continuous streams of data through sound is possible and has already been successfully implemented in the operating room. Anaesthetists already use a sonification of heart rate and oxygen saturation that are mapped to the changing speed and pitch of a stream of pulses (Bradakumar, Ball, Jefferson & Lindhoff, 2001). The present paper is concerned with extending operating room sonification to include six physiological variables.

There is little guidance about the process of designing a sonification (Kramer et al., 1999). The most comprehensive guidelines were developed by Barrass (1996) whose 'TaDa' process includes task analysis and rule based methods. However, there is very little information in the literature about how perceptual processes, the stimulus configuration, the data structure and expertise interact to affect design decisions, and very little about auditory attentional processes that might support the design of sonifications (Anderson & Sanderson, 2002). Information of this kind, however, will be crucial to designing successful sonifications. The research reported here is part of a larger three-phase program investigating these issues. The first phase deals with the perceptual discriminability of auditory

dimensions. This involves testing the discriminability of sound dimensions against chance, with zero, one and multiple distractor changes. The present research is part of this first phase. The second phase will investigate the effect of different mappings of similar information to different numbers of auditory streams. The third phase will investigate the effect of the sound signal structure on ability to monitor real physiological data. It is expected that physiological data is characterised by patterns and redundancies that will be used by domain experts to increase monitoring efficiency and accuracy.

1.1 Background to present research

Some recent studies have extended operating room sonification to include more than two physiological variables. Interestingly, all the reports in the literature of sonifications of more than two physiological variables have used two auditory streams. Loeb and Fitch (2000) designed a two-stream sonification with a cardiac stream and a respiratory stream. The cardiac stream conveys heart rate, saturation and blood pressure. The respiratory stream carries respiration rate, end-tidal CO₂ and tidal volume. Performance with this display on physiological monitoring tasks was contrasted with a visual only and a combined auditory and visual display. Accuracy and reaction time with the combined display was superior to either the visual or auditory display alone, suggesting that the auditory display was not sufficiently intelligible to support performance.

Seagull, Wickens and Loeb (2001) conducted similar comparisons and found that monitoring performance was poorest with an auditory display by itself. Performance on a secondary tracking task, however, was preserved in the auditory condition compared with the visual and combined conditions, suggesting that the auditory display allowed more attention to be directed to the secondary task. More efficient allocation of attention is a potential benefit of sonification and was clearly demonstrated in this study by the secondary task performance data. Unfortunately in Seagull et al. (2001) this benefit was offset by poorer performance in monitoring with the auditory display. The cause of this poorer performance is a question that remains unresolved.

A further example of a two-stream sonification was reported by Watson and Sanderson (1999; 2001). Anaesthetists performed a monitoring task equally well with a sonification, a visual display and a combined visual and auditory display. In addition, participants performed a secondary task involving simple arithmetic calculations concurrently with the monitoring task. Anaesthetists' performance on the secondary task was superior for the auditory display condition. The sonification allowed attention to be maintained on the secondary task while performance of the patient monitoring task was still high. The results for a sample of students who were untrained in physiology showed lower performance levels overall, and the sonification led to worse performance than the other conditions. The contrast between the performance of the anaesthetists and the student sample indicates the beneficial effects of expertise.

1.2 Unresolved issues

The above studies indicate when there are two auditory streams untrained participants have difficulty with an exclusively auditory monitoring task, compared to monitoring visual displays and combined displays. The cause of these difficulties has not been investigated. There are likely to be multiple factors affecting performance, including organization of the display, knowledge of the processes being monitored and the limits of human auditory processing. It is unclear how much information can be successfully monitored in the auditory modality and whether the perceptual organization of the auditory stimulus affects this. For example, information from six physiological variables could be organized into either one sound stream with six changing auditory dimensions, two sound streams with three changing auditory dimensions, (as described in the previous studies), or three sound streams each with two changing auditory dimensions. Would performance be any different with these different stream configurations, even though each configuration contained the same amount of physiological information, and how would expertise interact with stream configuration?

The studies reported here are part of a program of research that aims to answer these questions. In order to begin to address these problems, we explore the boundary condition: is it possible to create a single stream sonification with six changing auditory dimensions? The well-documented perceptual interactions of auditory parameters might limit the feasibility of a one-stream sonification (Boff &

Lincoln, 1986). We report the results of two studies, which were conducted consecutively. Study One was the result of work towards a successful one-stream six-dimensional sonification and was the final validation of a particular mapping. However, we were aware that some auditory dimensions produced better results than others. In particular, we were concerned with the effectiveness of amplitude, tremolo and width. In Study Two we therefore made adjustments to some of the mappings in order to determine whether the discriminability of the parameters could be improved. Since one of the questions of interest was whether discriminability improved from Study One to Study Two, combined results will be reported. The aims of the studies were (1) to establish six auditory dimensions with sufficient discriminability and range of values that could be used as the basis of a single stream sonification, and (2) to compare the discriminability of the mappings used in Study One and Study Two.

2 Method

The studies were both within-subjects designs with eleven participants in each. Participants were undergraduate psychology students who participated in the study for course credit. The experiment assessed ability to detect change in six auditory dimensions under three conditions: no other dimensions changing (no-distractor), one other dimension changing (no-distractor), five other dimensions changing (five-distractors). The stimuli were streams of pulses approximately ten seconds long (length was slightly variable to ensure that as speed changed whole pulses were heard). The six auditory parameters used were: amplitude, frequency, timbre, pulse speed, tremolo rate and pulse width. Participants monitored a target dimension for change and were asked whether the target dimension increased, decreased or stayed the same. Uncertainty was introduced by randomly varying the position of the change in the stream of pulses, always ensuring that at least two pulses were heard both before and after the change. All stimuli were constructed using Csound sound synthesis software and stored as .wav files. The stimuli were presented on a Pentium III personal computer and played through Philips digital speakers.

There were several differences between the stimuli used in Study One and Study Two and these are summarised in Table 1. The major differences are explained further here. First, ghosting was not used in Study One. Amplitude was reduced to zero between the pulses so that each pulse was separated by silence. Ghosting was used in Study Two. To produce ghosting the amplitude was not reduced to zero between the pulses, but was attenuated, producing a softer echo of the pulse between each pulse. This was designed to allow more time for listeners to extract other information about the sound, such as tremolo, from the echo when pulse width was short. Second, timbre was operationalised differently. In Study One timbre was operationalised as changes in harmonics produced by the changing centre frequency of a band pass filter. This produced a change from a full sound to a thin sound. In Study Two formant frequencies were used to produce a range of vowel sounds that formed a linear dimension with nine steps. The sound changed from an “AR” sound to an “OO” sound in nine steps. It was thought that formant frequencies, being more complex, would be less likely to be confused with the fundamental frequency and therefore less susceptible to interference from simultaneous pitch changes.

Table 1 Comparison of stimuli used in Study One and Study Two.

	Ghosting	Amplitude	Frequency	Timbre	Pulse speed	Tremolo	Pulse width
Study One	No ghosting – amplitude dropped to zero between pulses	Csound amplitude values used Nine values used	Frequency intervals determined by Herz	Changing centre frequency of band pass filter -full to thin sound	Pulses per second	Number of amplitude fluctuations within pulse	Duration of pulse in seconds
Study Two	Ghosting – amplitude reduced between pulses	Decibel values used Seven values used	Mel scale of subjective pitch used to ensure equal subjective distance between steps	Formant frequencies – “AR” to “OO” sound	Pulses per second	Amplitude fluctuations measured in cycles per second	Duration of pulse in seconds

The range of values for each dimension was based on the ranges of the physiological variables in the anaesthesia domain. Preliminary testing was used to indicate that the increments would be easily discriminable. All dimensions had to have good discrimination between steps. Thus, for all dimensions for which data on just noticeable differences were available, steps were more than one JND apart. For some parameters the JND had to be estimated and for others, such as tremolo rate there were no data available. From the full range of values for each parameter three values were sampled from the low, middle and high end of the range. These nine values formed the basis of testing. In each range there were two stimuli where the value of the target dimension increased, two where it decreased and three where it remained the same. This created 21 trials per dimension (7 trials in each of 3 ranges). For the one-distractor trials only the middle range was tested to ensure that the number of trials and therefore the length of the experiment remained feasible. For these trials each dimension was tested with every other dimension (5 combinations per dimension), creating 35 trials per dimension.

An explanation of the auditory dimensions was provided to participants and examples of changes in the target dimension were played prior to commencing the trials to orient attention to the target dimension. Altogether participants heard four examples of each dimension in different familiarisation sessions. The order of presentation of the dimensions was randomised within each condition and the order of presentation of the trials was randomised within each dimension. All participants received the no-distractor condition first, followed by the one-distractor and five-distractor conditions.

3 Results

3.1 Effect of number of distractors

Irrelevant changes in distractor dimensions reduced participants' ability to accurately detect changes in the target dimension. Performance was worse when many distractor dimensions changed compared to when one changed. A between-subjects three-way ANOVA comparing the results for Study One and Study Two showed no main effect for study $F(1,20)=0.79$, $MSE=.037$, ns, a main effect for parameter $F(5,100)=51.60$, $MSE=.014$, $p<.000001$ and a main effect for number of distractors $F(2,40)=218.05$, $MSE=.011$, $p<.000001$. There was also a significant interaction between study and parameter. A Newman-Keuls post hoc comparison showed that performance for amplitude was significantly higher in Study Two, while performance for formants in Study Two was significantly lower than performance for harmonics in Study One.

For Study One (see Figure 1(a)) a repeated measures ANOVA with the factors of parameter and number of distractors, showed a highly significant main effect for both parameter, $F(5,50)=21.071$, $MSE=.016$, $p<.000001$, and number of distractors, $F(2,20)=92.638$, $MSE=.014$, $p<.000001$, but no interaction. Results for Study Two are shown in Figure 1(b). In a repeated measures ANOVA there was a significant main effect for both parameter, $F(5,50)=43.99$, $MSE=.012$, $p<.000001$ and number of distractors, $F(2,20)=138.17$, $MSE=.009$, $p<.000001$ and a significant interaction between parameter and number of distractors, $F(10,100)=3.65$, $MSE=.009$ $p<.000359$.

In order to provide a stringent test of how good or bad performance was for each dimension, each was tested against the level of chance performance, $p(\text{correct/chance responding})=0.33$. The value for significance was corrected for the number of tests performed within each level of distraction, to preserve an overall alpha value (Type I error) of .05. In Study One, performance for all six parameters in the no-distractor condition was significantly better than chance, and in the one-distractor condition performance for all parameters except tremolo was significantly better than chance. Since this was a very conservative test of performance in a difficult task, we also report marginally significant results. At $0.1 > p > 0.05$, performance for tremolo in the one-distractor condition was also significantly better than chance. For the five-distractor condition performance dropped to chance level for all parameters.

In Study Two performance for all six parameters was significantly better than chance in the no-distractor condition. In the one-distractor condition performance for amplitude, frequency and speed only was significantly better than chance. Performance for formants, tremolo and width had dropped to chance level with only one-distractor change. For the five-distractor condition performance dropped to chance level for frequency, formants, tremolo and width. However performance for speed remained

significantly better than chance, while performance for amplitude was marginally significantly better than chance.

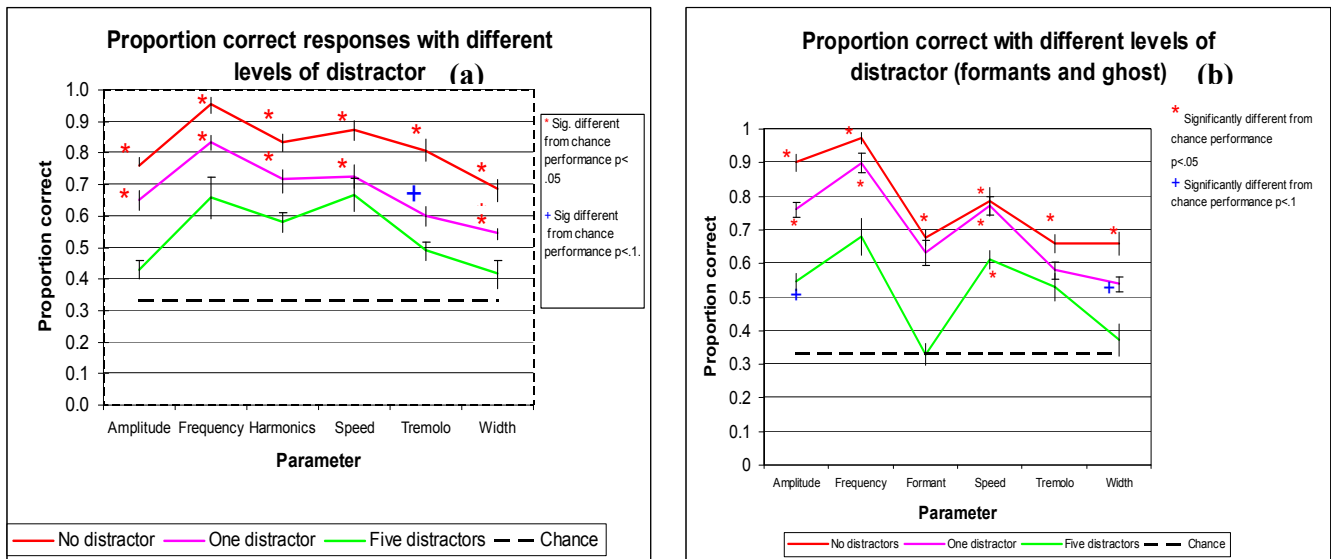


Figure 1: Proportion correct responses with different levels of distractor, showing performance in relation to chance.

3.2 Interactions between dimensions

The effect of simultaneous changes in dimensions was of interest to determine the degree of interference that would occur. The effectiveness of most dimensions was severely reduced by interactions.

- Amplitude.** Figure 2a shows that for Study One detection of amplitude changes was reduced to chance level for all combinations of amplitude and distractor parameters. Performance for amplitude was better in Study Two, confirming our expectations that reducing the number of steps and converting values to decibels could improve amplitude performance. Performance for all combinations of amplitude and distractor parameters, except amplitude and frequency, was significantly better than chance.
- Frequency.** The results for detection of changes in frequency were very similar for Study One and Study Two, as shown in Figure 2b. Performance for frequency was not significantly different from chance when frequency change detections were accompanied by changes in amplitude and harmonics/formants.
- Harmonics/Formants.** Detection performance for harmonics/formants dropped to chance level for both studies when changes simultaneously occurred in amplitude and frequency and width. Contrary to expectations that formants would be less vulnerable to interference from frequency, performance for harmonics was severely reduced by changes in frequency (see Figure 2c).
- Pulse speed.** Figure 2d shows the results for speed, indicating that for both Study One and Study Two, performance in detecting speed changes was at chance level when simultaneous changes also occurred in tremolo. For other combinations of speed and distractor parameters performance was either significantly better than chance or marginally different from chance. The presence of the ghosting in Study Two improved performance for the speed/tremolo combination above the level of performance in Study One.

- **Tremolo.** From Figure 2e, performance for tremolo was significantly better than chance in Study One when tremolo changes occurred in combination with changes in speed, and also in Study Two when tremolo changes were accompanied by changes in width. The presence of ghosting in Study Two resulted in a dramatic improvement of performance for the tremolo/width combination, confirming expectations that ghosting extended the period of time available to extract information about the sound, compared to a pulse stream with no ghosting.
- **Pulse width.** Finally, for detection performance for width, no combinations were significantly different from chance. Figure 2f shows that for Study Two, performance for the width/tremolo combination was higher, suggesting that ghosting also reduced interference from tremolo when width judgements were being made.

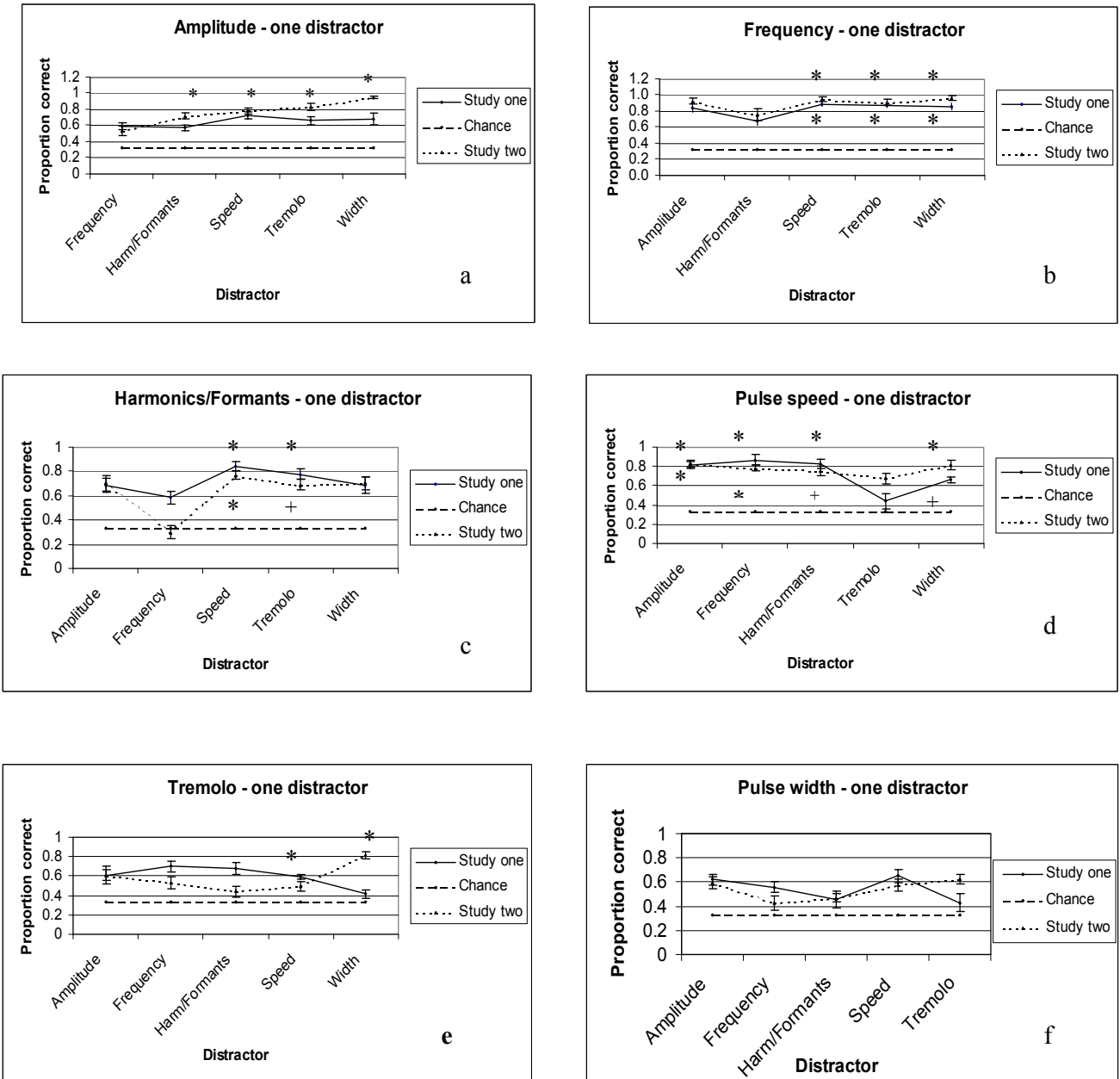


Figure 3: Proportion correct responses with different levels of distractor, showing performance in relation to chance

4 Discussion

The first aim of the studies reported here was to investigate whether it is feasible to create an intelligible single-stream sonification in which information is carried on six auditory dimensions. This was to provide a basis for examining selective and divided attention with different stream configurations. Results indicated differences in the effectiveness of different dimensions and also differences based on the amount of distraction experienced. Attempts to reduce acoustic interactions between parameters by reconfiguring the stimuli had mixed success. Although the results appear to indicate that some of the dimensions used were not strong, it must be noted that the discrimination task was difficult. With no contextual information, participants were asked to detect a change of the magnitude of a single increment with uncertainty about when the change was going to occur. They also had to detect the direction of the change. Combined with the requirement that performance should be better than chance responding, this represents a stringent test of the one-stream, six dimension sonification. In the anaesthesia domain it is likely that a sonification would be used to monitor trends occurring in the physiological variables. A trend would involve change over several increments rather than just one. However it is also possible that in some critical situations a change of one increment in a variable would need to be detected, necessitating a stringent test of discriminability.

Performance for all parameters was significantly better than chance in the no-distractor condition in both studies. Allowing marginally significant results, in Study One performance for all parameters was also significantly better than chance in the one-distractor condition. However, for Study Two only amplitude, speed and width performance were significantly different from chance. Performance was no better than chance for all dimensions in the five-distractor condition for Study One, but amplitude and speed performance were significantly better than chance in Study Two. Clearly there are limits to detecting change on auditory dimensions when changes on many other dimensions are also occurring. However, in the anaesthesia domain it would be rare for all six of the variables to change simultaneously and in arbitrary directions. The absence of context, as would be provided by real physiological data would also have contributed to the difficulty of the task.

The second aim of the studies was to compare the different stimulus configurations to determine which was the most successful mapping. The data from Study One indicated that some combinations of dimensions interacted to produce interference that significantly reduced the discriminability of both dimensions. Interference occurred between dimensions where judgements of the same percept were required. For example, interference between frequency and harmonics occurred because timbre was operationalised as the selection of harmonics in a particular frequency range, leading to confusion about the frequency of the fundamental. Similarly, tremolo was operationalised as the speed of amplitude fluctuations within the pulse and led to confusion with the speed of the pulse stream itself. The interference between tremolo and width occurred for different reasons. When pulse width was short, there may not have been enough time to gauge the speed of the tremolo. Tremolo might also have led to some uncertainty about the onset and offset of the pulse, which is crucial in judging pulse width. For Study Two an attempt was made to develop sounds that minimised these interactions. Ghosting was used to enable more information to be extracted about the sound than would normally be perceivable in a short duration pulse. Timbre was operationalised using formant frequencies in an attempt to reduce interference between harmonics and frequency.

The changes had mixed success. First, the formant parameter was significantly less discriminable than harmonics and much more vulnerable to interference from frequency. This was surprising. Participants found it very difficult to accurately identify the direction of a change even though they had heard a change, suggesting that there were problems with the linearity of the dimension. Second, ghosting was also a mixed success. While there was a performance improvement for particular target and distractor combinations, such as speed/tremolo and tremolo/width, overall performance appeared to be reduced by the ghosting. A further observation was that some listeners could sometimes perceive the ghosting as two sounds, rather than one sound reducing in amplitude between the pulses. For some listeners there appeared to be a soft continuous sound that continued behind the louder pulse, which was particularly noticeable when this effect had been pointed out to them. Stream segregation by loudness is possible and for listeners who are trying to hear separate soft and loud streams the loudness difference needs to be only 3-4 decibels to produce this effect. For listeners who are trying to hear one stream the required loudness difference is much greater (Bregman, 1990). The next step in the planned research is to compare this sonification with a two stream and three stream sonification and the

possibility that ghosting can make the single stream segregate into two streams introduces an unwanted possible confound into the design of future experiments.

The results from this experiment have led to several conclusions. First, we know that the mappings used in Study One were the most successful. We plan to retain the mel and decibel measurements for frequency and amplitude and reduce amplitude, tremolo and width to seven steps in order to increase discriminability. Second, the next stage in this program of research is to develop two stream and three stream configurations of the same dimensions in order to examine selective and divided attention with different stream configurations. One of the major advantages of dividing the stimulus into a number of streams is the potential for reducing interference between dimensions. Dimensions will interact within a stream, but if stream segregation were successful interaction would not occur across streams. The results of these studies indicate that for the two stream and three stream sonifications it is crucial to separate frequency and harmonics, speed and tremolo, and tremolo and width.

5 Design implications

The research reported here is part of a larger program that addresses the problem of how to design an effective sonification for the anaesthesia domain. The larger program will address the perceptual discriminability of auditory dimensions compared to chance performance, the operation of selective and divided attention with different numbers of auditory streams and different dimensional mappings and the effect of expertise. The perceptual relationships and interactions are likely to change when physiological data streams are used because changes will be auto-correlated. In the present study participants were largely untrained and were simply instructed to listen for a change in the sounds. The tasks were not contextualised and the natural patterns and redundancies that would occur in real physiological data did not exist. Further work will investigate whether the relationships found in this study generalise to domain experts (anaesthetists). We are thus advocating a combined approach to sonification design based on an understanding of perceptual processes, data structure, and stimulus configuration and also an understanding of the interaction of those processes with domain knowledge.

6 References

- Anderson, J. & Sanderson, P. (2002). Sonification for real-time processes: Issues and a demonstration. Paper submitted for publication in *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society*, Baltimore, MD. September 30-October 4, 2002.
- Barras, S (1996). TaDa! Demonstrations of Auditory Information Design. *Proceedings of the International Conference on Auditory Displays (ICAD96)*, Palo Alto, California.
- Boff, K. & Lincoln, J. (1986). Engineering data compendium. Human perception and performance. Volume 1. Wright-Patterson A.F.B., Ohio: Harry G. Armstrong Aerospace Medical Research Laboratory.
- Bradrakumar, A., Ball, D.R., Jefferson, P. & Lindhoff, G. (2001). Pulse oximetry – to bleep or not to bleep? (Letter to the editor). *Anaesthesia*, 56, 184.
- Bregman, A. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.
- Gaba, D.M. & Howard, S.K. (1995). Situation awareness in anesthesiology. *Human Factors*, 37, 20-31.
- Kramer, G. et al. (1999). *Sonification report: Status of the field and research agenda*. Prepared for the National Science Foundation by members of the International Community for Auditory Display.
- Loeb, R.G. & Fitch, T. (2000). Laboratory evaluation of an auditory display designed to enhance intraoperative monitoring. *The Society for Technology in Anesthesia annual meeting 13-15 January 2000 Orlando*. Abstract from anestech.org/publications. File: Annual_2000/Loeb.html
- Seagull, F.J., Wickens, C.D. & Loeb, R.G. (2001) When is less more? Attention and workload in auditory, visual, and redundant patient-monitoring conditions. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, 2001.
- Watson, M., Russell, W.J. & Sanderson, P.M. (1999). Ecological interface design for anesthesia monitoring. *Proceedings of the Australian/New Zealand Conference on Computer-Human Interaction (OzCHI99)*. (pp. 78-84). Wagga Wagga: Charles Sturt University.
- Watson, M. & Sanderson, P. (2001). Intelligibility of sonifications for respiratory monitoring in anesthesia. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, 2001.