# Evaluation of Psychoacoustic Sound Parameters for Sonification

Jamie Ferguson
University of Glasgow
Glasgow, Scotland
j.ferguson.4@research.gla.ac.uk

Stephen A. Brewster
University of Glasgow
Glasgow, Scotland
stephen.brewster@glasgow.ac.uk

## ABSTRACT

Sonification designers have little theory or experimental evidence to guide the design of data-to-sound mappings. Many mappings use acoustic representations of data values which do not correspond with the listener's perception of how that data value should sound during sonification. This research evaluates data-to-sound mappings that are based on psychoacoustic sensations, in an attempt to move towards using data-to-sound mappings that are aligned with the listener's perception of the data value's auditory connotations. Multiple psychoacoustic parameters were evaluated over two experiments, which were designed in the context of a domain-specific problem - detecting the level of focus of an astronomical image through auditory display. Recommendations for designing sonification systems with psychoacoustic sound parameters are presented based on our results.

## CCS CONCEPTS

• **Human-centered computing → Auditory feedback**;

## KEYWORDS

Sonification, Auditory Display, Psychoacoustics

**ACM Reference Format:**
Jamie Ferguson and Stephen A. Brewster. 2017. Evaluation of Psychoacoustic Sound Parameters for Sonification. In *Proceedings of 19th ACM International Conference on Multimodal Interaction (ICMI'17).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3136755.3136783

## 1 INTRODUCTION

The auditory sense is an important factor in human-computer interaction and data visualisation, and as large and complex multidimensional data sets become more common, there is a growing need for as many perceptual and technological resources as possible to be utilised to display data effectively. Using solely visual techniques fails to utilise the human auditory system's powerful anomaly and pattern recognising potential and makes data inaccessible to people with visual impairments.

*Sonification* is "the use of non-speech audio to convey information" [15] and more specifically, data sonification is "the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation" [15]. There are a number of examples of successful applications of sonification from many fields including astrophysics [5], neurology [25], palaeontology [7] and computer accessibility [14].

When designing a sonification system, data values are used to manipulate an acoustic parameter, such as frequency (pitch) or tempo. Here, the most important design decision is choosing the optimum mapping of data values to acoustic parameters. An effective sonification design relies on a compromise between "intuitive, pleasant and precise" sonic characteristics [12]. However, there has been little research conducted to determine the range of acoustic parameters that are most effective at conveying particular data values and there is minimal theory or experimental evidence to advise sonification researchers and designers.

23.8 % of data-to-sound mappings reported in sonification studies use pitch as an acoustic parameter [6]. The human auditory system has a high resolution of pitch [19] and it is almost certainly due to both this high resolution, and the fact that pitch has an inherent association with music, that it has been used so extensively as a sonification parameter. However, the need to thoroughly investigate other acoustic parameters for use in sonification is twofold: certain contexts and data values may require an acoustic mapping that is aligned with the user's preconceived perceptions of the context or data value being sonified; and in multidimensional sonifications, a number of data-to-sound mappings may be required, therefore if pitch is chosen to convey one data type, other acoustic parameters will be needed to convey the rest of the data set.

Taking guidance from psychoacoustic research when designing sonification mappings may provide data-to-sound mappings that are aligned with perceptions of the data type being sonified. For example, many people would describe the sounds relating to a thunderstorm to be a low, rumbling and booming sound, therefore in some contexts, using the pitch of a note to sonify data relating to thunderstorm intensity may be conflicting with the listener's perception of what a thunderstorm sounds like. Whereas, in this case, a more perceptually aligned mapping could implement the psychoacoustic sensation of *booming* (the presence of low frequencies in a sound) [11].

In this paper we discuss the potential of designing acoustic parameters for sonification based on psychoacoustic sensations and present a real-world application. Our research has the following aims: to evaluate the feasibility of psychoacoustic parameters for use in a sonification task; to assess the resolution of these acoustic parameters; to guide the design of future sonification systems that implement these parameters, and to investigate the accuracy of these parameters in a sonification task compared to a visual equivalent.

## 2 RELATED WORK

### 2.1 Psychoacoustics

Psychoacoustics - "the science of the hearing system as a receiver of acoustical information" [30], aims to model the complexity of auditory sensation in physical parameters. Zwicker and Fastl's work [30] collects theoretical and experimental research that presents algorithms which physically quantify auditory sensations such as loudness, roughness and sharpness. Application of modelled psychoacoustic sensations is common in industrial acoustics design to optimise the acoustic qualities of products and appliances [17, 22, 23]. Psychoacoustic models applied in this context are often defined as *sound quality metrics* . Sound quality in this context is defined as "a perceptual reaction to the sound of a product that reflects the listener's reaction to how acceptable the sound of that product is" [16]. Psychoacoustic parameters are an important factor in product design, as the aesthetic and cognitive effects of a product's sound emission may be important to a user's experience with a product. They are not commonly used in sonification design.

### 2.2 Perception of Auditory Displays

Experiments conducted by Walker and Kramer [27] showed that data-to-sound mappings that seem effective at first glance may in fact result in reduced performance, as the listener may find the mappings perceptually confusing or unpleasant. Further studies by Walker [26] applied experimental methods used in psychophysics to assess what acoustic parameter may be most effective for a given data value. These studies showed the need to investigate the perceptual reactions of users to a given data-to-sound mapping, to allow the most effective design possible. Neuhoff and Heller [18] proposed a method of auditory display that uses data-to-sound mappings that utilise a "pre-existing cognitive structure in which to interpret the sonified changes in a variable". This method of auditory display was motivated by the fact that when a listener is evaluating some fundamental acoustic dimensions such as pitch, they can encounter interference from other such dimensions like loudness, as the two are perceptually integrated [10]. Peres and Lane [21] conducted experiments which utilised these integrated acoustic dimensions in an auditory display. Ferguson et al. [9] proposed the use of psychoacoustic models for sonifcation and discussed the potential benefits, however no experimental assessment was conducted.

## 3 PSYCHOACOUSTIC PARAMETERS

In this section, we outline the theory of three fundamental psychoacoustic parameters and discuss their potential benefits for use as an acoustic parameter in an auditory display.

### 3.1 Fluctuation Strength & Roughness

Fluctuation strength quantifies the subjective perception of slower amplitude or frequency modulation of a sound (modulation frequencies of up to 20 Hz, reaching its maximum near 4 Hz) [8]. Roughness quantifies the subjective perception of rapid amplitude or frequency modulation of a sound (modulation frequencies from around 15 Hz to 300 Hz, reaching its maximum near 70 Hz) [30]. The same rapid amplitude modulation that elicits the psychoacoustic sensation of roughness is also used to explain the acoustic component of dissonance [13].

### 3.2 Sharpness

Sharpness quantifies the high-frequency content of a sound - a larger proportion of higher frequencies equates to a sharper sound [30].

### 3.3 Sonification Potential

Due to the above parameters' ability to describe the quality of a sound, it is possible that their descriptive qualities could effectively convey the quality of another dimension. Roughness and sharpness are multimodal descriptions of texture or quality as they apply in the visual and haptic modalities as well as the auditory. Therefore, the cross-modal nature of these parameters, in addition to the importance of roughness to the perception of dissonance may offer novel data-to-sound mappings that are more aligned with perceptions of the data parameter being sonified. This paper investigates this hypothesis.

## 4 APPLICATION

A domain-specific problem in which these ideas may be implemented was developed with the Innovators Developing Accessible Tools for Astronomy (IDATA) research initiative [1]. IDATA uses astronomy to develop transferable skills in computational thinking, software design and computer accessibility in both sighted and visually impaired students. Students taking part can access the Skynet Robotic Telescope Network [2] to remotely take images of astronomical objects using a large telescope. They can then analyse and share these images using the Skynet Junior Scholars portal and its related analysis software [3]. One of the main goals of the project is to provide a platform where visually impaired students can carry out these tasks unaided by a sighted person. Furthermore, IDATA aims to ensure that the accessibility tools they provide for visually impaired students are as equal to the tools available to sighted students in terms of functionality and accuracy.

The application described in this section was formed during focus-groups and interviews with IDATA staff to gather information on their user group, domain-specific problems and how sonification may provide a solution to them. One of the first actions after taking an image via a telescope is to judge the overall quality of the image, assessing if it is in-focus and if the tracking was set correctly. IDATA reported that this fundamental step is currently impossible for visually impaired students to carry out without the aid of a sighted person. IDATA expressed an interest in attempting to solve this problem using an auditory solution, therefore a sonification-based approach was formulated in which the acoustic parameters to display the level of focus of an image were designed based on psychoacoustic sensations. As discussed above, the connotations of quality and texture present in psychoacoustic parameters may provide a data-to-sound mapping for this application that is more aligned with a user's preconceived perceptions of image focus.

## 5 EXPERIMENT 1

An experiment was conducted to provide an initial evaluation of utilising these psychoacoustic parameters as acoustic parameters in a data-to-sound mapping. The experiment was designed with the application discussed above in mind and we planned the experiment with input from the IDATA team. The aim of the experiment was to identify how accurately participants could determine specific points on a 10-point scale for each sound parameter, thus giving an

[1] http://www.yerkesoutreach.org/idata.html
[2] https://skynet.unc.edu/
[3] https://skynetjuniorscholars.org/

initial overview of the acuity of each psychoacoustic parameter in a sonification context. This 10-point scale was framed in terms of the image focus application mentioned previously and throughout the experiment, the acoustic parameters were discussed as being "mapped to" or "representing" the level of focus of an image.

## 5.1 Participants

21 participants took part in the study (12 female, 9 male; mean age = 24.3 years, SD of age = 7.7 years). All participants were university students and staff. As this study was an initial investigation to evaluate the effectiveness of these sound parameters, only sighted participants took part. The results from this experiment will influence future studies focusing on visually impaired people's interactions with these sound parameters. All participants reported no uncorrected vision impairment, no hearing impairment and no music/sound related neurological condition such as amusia [20].

## 5.2 Design

Four conditions were examined in which the single independent variable in each condition was the acoustic parameter used to represent the level of focus of an image. Roughness, sharpness and a combination of both were chosen, with the latter used to investigate if a combination of sound qualities is better than each alone. In addition, pitch was included as a condition, to provide a comparison to a more traditional data mapping. The experiment used a within-subjects design. Dependent variables collected during the experiment included: number of correct responses, reaction time for each response and perceived polarity of data-to-acoustic parameter mapping.

During the experiment, data were collected on the number of correct responses to the level of focus being sonified. Percentage correct scores were calculated for totally correct responses (0 % error) and for responses that were 1 level away from the correct response (10 % error). IDATA reported that it would be useful to evaluate what resolution is most accurate in this task, therefore since there were 10 levels of each parameter, (0 % error) represented a resolution of 10 levels and (10 % error) represented a resolution of 5 levels.

## 5.3 Stimuli

Ten stimuli were used for each condition. Each stimulus was 2 seconds in length, this was chosen as work by Brewster [3] showed that information encoded in sounds can be obtained from a stimulus between 1 and 2 seconds in length. Each stimulus had an amplitude envelope with a 0.2 second linear ramp onset (attack) and offset (release). An amplitude envelope was included in the sound design, as an abrupt start or stop of a sound can be perceived as unpleasant [1]. All stimuli were created in the Supercollider programming language [4]. The acoustic design of these stimuli is described below.

*5.3.1 Roughness.* 100 % sinusoidally amplitude modulated broadband noise with modulation frequencies of 14, 17, 21, 25, 30, 36, 43, 52, 62 and 75 Hz. This range of modulation frequencies was chosen based on Fastl & Zwicker's estimation that, in the case of amplitude-modulated pure tones, an increment in roughness is perceptible in increments of 17 %.

Fastl & Zwicker's threshold is based on increments of modulation degree not modulation frequency and for pure tones not broadband noise. In this experiment we used broadband noise as opposed to pure tones as the carrier signal for roughness. This was because the sharpness condition requires the use of a complex noise signal, therefore noise was used in the roughness condition to mitigate any performance differences based on the carrier signal type. Modulation frequency was chosen as the dependency of roughness to vary instead of degree of modulation, as modulation frequency provides a larger range of potential levels of roughness.

Informal pre-testing was carried out to determine an approximate number of roughness increments that could be perceived for corresponding increments in modulation frequency. These tests suggested that between modulation frequencies of 14 and 75 Hz, differences in roughness are just-noticeable in steps of around 17 %.

*5.3.2 Sharpness.* Critical-band wide white noise. For a given frequency, a critical-band is the smallest range of frequencies around it that activate the same area of the basilar membrane - the primary sensory apparatus of the ear [24]. There are 24 critical bands, this scale being the Bark scale [29]. Stimuli in this condition were at Bark levels 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20. This range of critical-bands was chosen based on informal pre-testing which suggested that an increment in sharpness was perceptible per increment in Bark level. As increments of 1 Bark in the range of 24 Bark were perceptible and a resolution of only 10 was required for this study, we decided to use increments of 2 Bark. The range began from 2 Bark, as some pre-test participants reported difficulty in hearing 1 Bark. Loudness for each stimulus was normalised using Supercollider's built in library for basic psychoacoustic amplitude compensation - AmpComp.

*5.3.3 Combined Roughness & Sharpness.* Direct pairing of corresponding roughness and sharpness stimuli. This condition was included to assess if a combined sound quality parameter mapping was more accurate than its individual parameters.

*5.3.4 Pitch.* pure tone with frequencies of 100, 200, 400, 800, 1000, 1400, 1800, 2400 and 3200 Hz. This frequency range and the choice of using a pure tone was based on the stimuli used by Walker [26]. Loudness for each stimulus was compensated.

## 5.4 Procedure

The experiment consisted of 4 blocks of 30 trials (1 block for each condition). In each trial, participants were presented with a stimulus which lasted for 2 seconds and after a 2 second pause, was repeated. Based on this stimulus, participants were asked to respond on a 10-point scale, how clear and in-focus they believed they image represented by the sound was (10 being a perfectly clear image, 1 being very blurry and out of focus). They indicated their response via a mouse, on a scale shown on-screen. Each stimulus was presented 3 times, with the order of presentation randomised and the order in which each block was presented was also randomised.

Before starting the experiment, participants were trained to use the interface by performing 2 mock trials. They received no feedback on their performance during this training, as part of the experiment's aim was to investigate how participants perceived the psychoacoustic sound parameters in an auditory display context.

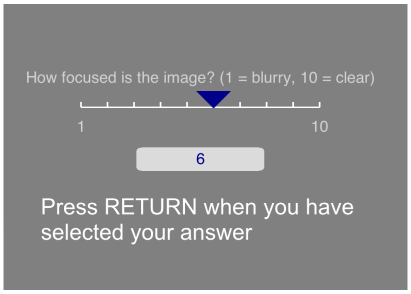


**Figure 1: The focus-level response dialogue.**

Furthermore, participants received no training on the acoustic definitions of any of the parameters that were used.

*5.4.1 Polarity Choice.* At the beginning of each block, participants were presented with the 10-step scale of each acoustic parameter in full 3 times. After hearing all 10 stimuli for a given parameter, the participant indicated on an on-screen interface either *start* - if they perceived the first sound in the scale to represent a clear image or *end* - if the perceived the final sound to represent clarity. Participant's responses were inverted for our analysis if they selected *start*, as this indicated they perceived the scale the other way around (i.e. ascending pitch = descending focus). By inverted we mean that if a stimulus of focus = 0 is presented and the response = 10, then response = 10 - participant response. All analysis was performed after these inversions had been carried out.

## 5.5 Results

Three participants were eliminated from the evaluation as they confused their polarity choices on all conditions - meaning that in each condition they deviated from the polarity choice they initially indicated. The mean Pearson correlation was calculated between each sound level and the participant's response at that level. The criteria for elimination was: If after all necessary polarities were inverted, a participant's mean Pearson correlation over all levels showed a negative correlation, this indicated a significant diversion from their selected polarity for all conditions and they were eliminated from the evaluation. Therefore, the results reported are for the remaining 18 participants.

The results for responses in the 0 % error threshold (0 levels from correct response) showed an overall correct response rate of 20 %, with the pitch condition having the highest correct response rate of 28 % and sharpness having the lowest of 15 %. A one-factor analysis of variance (ANOVA) showed a main effect ($F(3,68) = 6.80$, $p = 0.009$). Of the psychoacoustic parameters, the combined condition showed the highest correct response rate of 19 %. *Post hoc* Tukey HSD tests showed that the pitch results were significantly higher than the roughness ($p = 0.03$) and sharpness conditions ($p = 0.009$). No other significant pairwise comparisons were found.

The results for responses in the 10 % error threshold (+/- 1 level from correct response) showed an overall correct response rate of 54 % and also showed the pitch condition to have the highest correct response rate of 67 %. The rest of the sound parameters had similar results to each other (Roughness = 50 %, Sharpness = 50 % and Combined = 49 %). A one-factor ANOVA showed a main

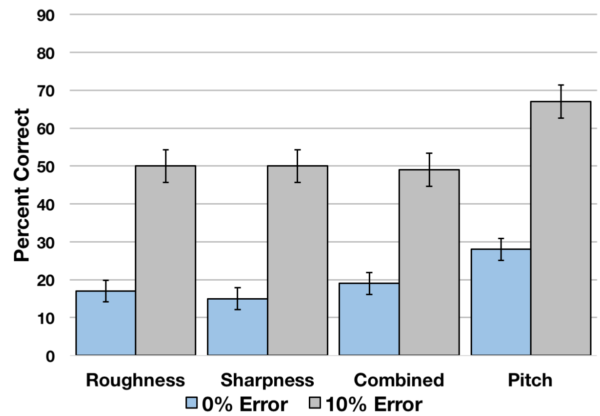


**Figure 2: Mean percentage of correct responses for each condition. Error bars show standard error.**

effect ($F(3,60) = 2.8$, $p = 0.04$). A *post hoc* Tukey HSD test showed no significant pairwise comparisons.

The mean number of correct responses per level was analysed to establish if more errors occurred at certain levels on the 1-10 scale. This analysis showed that for both 0 % error and 10 % error, all of the novel parameters show a decrease in correct responses at the higher levels of focus and a higher number of correct responses at the lower levels of focus (Figure 3).

## 5.6 Discussion

Although results overall were poor for responses with 0 % error, results for 10 % error suggest that focus can be detected through sonification more than two times more accurately with 10 % error as opposed to 0 % error. This indicates that a resolution of 5 levels rather than 10 may be more appropriate for this task, however more training may improve accuracy at a higher resolution. As was mentioned earlier, IDATA reported that they have no concrete requirement for the resolution of focus detection through sonification, therefore a resolution of 5 levels may be effective in their application.

The declining mean number of correct responses towards focus level 10 showed in the results for the psychoacoustic parameters (Figure 3), suggests that participants were successful at detecting if an image was significantly out of focus, but found difficulty in detecting if the image was in focus. This could be because none of the psychoacoustic sounds converged on a clear sound, i.e. even at focus level 10, all of the parameters' carrier sound was still white noise, therefore this could have confused participants. The pitch condition did not show this decline and it was the only parameter that wasn't based on white noise, further suggesting that this may be the reason for the decline in correct responses for the other parameters.

# 6 EXPERIMENT 2

Results from experiment 1 indicated that the sound design of the psychoacoustic sound parameters was not the most effective and

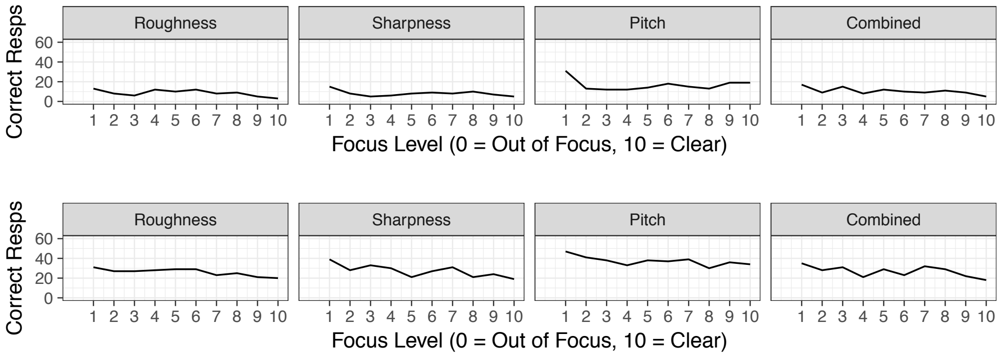


**Figure 3: Mean number of correct responses at each level, for 0 % error (top) and 10 % error (bottom).**

could potentially be improved. Therefore, a second experiment was conducted under the same experimental design and procedure, with a new set of psychoacoustic sound parameters. Furthermore, as IDATA aims to ensure that the accessibility tools they provide are equal to the tools available to sighted students, a visual condition was included. This was to evaluate how effective auditory analysis of image focus compared to visual.

## 6.1 Participants

20 new participants took part in Experiment 2 (12 female, 8 male; mean age = 23 years, SD of age = 3.4 years). Participants were recruited under the same criteria as experiment 1.

## 6.2 Design

Five conditions were examined in which the single independent variable was the parameter used to represent the level of focus of an image (four acoustic parameters and one visual parameter). Roughness (of a pure tone), noise, combined roughness & noise and pitch were chosen as auditory conditions. In this experiment, the pitch range was based on a musical scale, to investigate if it was more effective than the frequencies chosen in Experiment 1. The visual condition consisted of an astronomical image being artificially blurred to various degrees. During the experiment, the same data was collected as in experiment 1.

## 6.3 Stimuli

Auditory stimuli were designed in the same fashion as experiment 1 and the visual stimuli were artificially blurred using OpenCV [5]. The design of these stimuli is described below.

*6.3.1 Roughness.* 100 % sinusoidally amplitude modulated 1000 Hz pure-tone with modulation frequencies of 0, 2, 4, 7, 11, 16, 23, 34, 49 and 70 Hz. The roughness condition in the previous experiment showed declining performance toward focus level 10, this may have been caused by the range of roughness not converging on a clean sound. Therefore, in designing a range of roughness sounds for this experiment, we decided to use a pure tone as a carrier. The use of noise as a carrier in the previous experiment may have conflicted with participants' perceptions of a clear and in-focus image, as noise may be perceived as out-of-focus. As the transition between fluctuation strength and roughness is considered to be smooth rather than a strict border [30], we included the lower modulation frequency ranges of fluctuation strength. We included these lower modulation frequencies down to an unmodulated pure tone so that the range of roughness converged on a pure tone. The step size of this range was based on the same informal pre-testing as the roughness condition in Experiment 1, however the step size was increased, due to the larger range of available frequencies afforded by including the range of fluctuation strength.

*6.3.2 Noise.* This condition consisted of a 1000 Hz pure tone and broadband white noise, with the pure tone beginning at 100 % amplitude and noise at 0 %, respectively declining and ascending in 10 % increments. The aim of this condition was to utilise the effect suggested by the results of experiment 1 - that noise may be associated with blurriness or lack of clarity.

*6.3.3 Combined Roughness & Noise.* Direct pairing of corresponding roughness and noise stimuli. As the combined condition in the previous experiment showed slightly increased performance over the individual parameters, we included a combination in this experiment.

*6.3.4 Pitch.* pure sinusoidal tones in a C-major scale (plus 2 extra notes to make the range 10 notes long) beginning at middle C (C4, freq = 261.63 Hz) and ending on E6 (freq = 1318.51 Hz). Loudness for each stimulus was compensated.

*6.3.5 Visual.* 1700 x 2000 pixel, grey scale image of the dwarf galaxy M110 (Figure 4) taken on a telescope typical of the kind used in the IDATA project. The image was artificially blurred using OpenCV 's *blur* function. This takes as input a window (kernel) size in pixels and each pixel in the output is the mean of all of the pixels in the window around the corresponding pixel in the input [2]. The kernel sizes used were 0 (image with no blur), 10, 20, 40, 70, 110, 160, 220 and 440 pixels squared. The experimental procedure was identical to Experiment 1. For the visual condition, the stimuli were presented for the same amount of time as in the auditory conditions.

[5]http://opencv.org

**Figure 4: Image of M110 used in the visual condition (NOAO/AURA/NSF).**

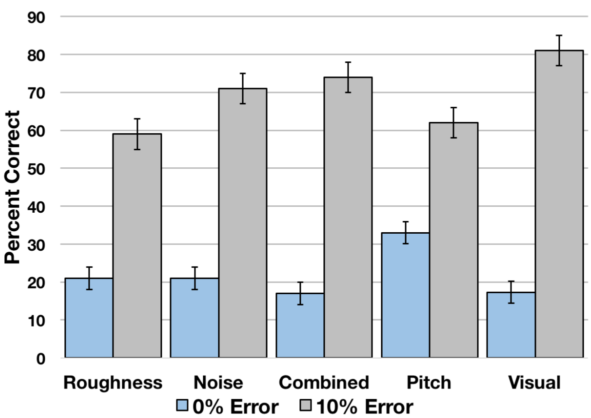


**Figure 5: Mean percentage of correct responses for each condition. Error bars show standard error.**

## 6.4 Results

The results for responses in the 0 % error threshold showed an overall correct response rate of 22 % (audio conditions only = 23 %), with the pitch condition having the highest correct response rate of 33 % with the combined and visual conditions having the lowest of 17 %. A one-factor ANOVA showed a significant main effect ($F (4,95) = 3.86, p = 0.006$). *Post hoc* Tukey HSD tests showed that the pitch results were significantly higher than the combined ($p = 0.03$) and visual conditions ($p =0.01$). No other significant pairwise comparisons were found.

The results for responses in the 10 % error threshold (+/- 1 level from correct response) showed an overall correct response rate of 69 % (audio conditions only = 66 %) and also showed the visual condition to have the highest correct response rate of 81 %. Of the auditory conditions, the combined condition performed best with a correct response rate of 74 % and the roughness condition showed the lowest with 59 %. An ANOVA showed a significant main effect ($F (3,60) = 4.35, p < 0.001$). *Post hoc* Tukey HSD tests showed that the visual condition performed significantly better than the pitch and roughness conditions. No other significant pairwise comparisons were found.

Analysis of correct responses per level showed that for responses in the 0 % error threshold, all conditions aside from pitch generally declined toward either end of the range of focus (Figure 6 - top). Responses within the 10 % error threshold (Figure 6 - bottom) indicated that the pitch and roughness conditions showed a decline at higher levels, whereas the noise, combined and visual conditions showed an increase at either end of the range of focus.

As both experiments were conducted under the same conditions, an ANOVA was carried out on both sets of experiment conditions together. This showed a significant main effect ($F (8, 155) = 45.46, p < 0.001$). *Post hoc* Tukey HSD tests showed that all conditions in experiment 2 performed significantly better than those in experiment 1 ($p < 0.001$).

## 6.5 Discussion

As in experiment 1, results for responses within 10 % error showed more than twice the number of correct responses than responses within 0 % error. The declining mean number of correct responses within 0 % error at the extremes of the range of focus shown in Figure 6 (top), suggests that for all conditions with the exception of pitch, participants found difficulty in detecting focus level at the extremes of the range during these conditions. However, Figure 6 (bottom) indicates that all conditions with the exception of pitch and roughness show an increase at the extremes of the range of focus. This suggests that in the noise, combined and visual conditions, participants were more successful at detecting focus level at the extremes of the range within 10 % error, as opposed to detecting the level with no error. Combined, the results shown in Figure 6 suggest that a resolution of 5 levels is more effective than 10 for the psychoacoustic parameters used.

## 6.6 Design Recommendations

Based on our findings, we suggest the following recommendations to researchers and designers creating sonification systems and auditory interfaces involving psychoacoustic sounds.

- For applications such as the focus detection application presented in this paper, acoustic parameters that converge on a clear, pure sound are better. All of the psychoacoustic sound parameters used in experiment 2 converged on a pure tone and these sounds performed significantly better than the parameters used in experiment 1. This is significant, as this finding suggests that listeners associate noise and roughness with an undesirable attribute such as blur in this case and a clear tone with the desirable quality of clarity and focus.

- Listeners can detect the data value being sonified through psychoacoustic sounds better at a resolution of 5, rather than 10. In both experiments, there were more than twice as many correct responses within 10 % error than there were within 0 % error. This suggests a resolution of 5 levels may be more accurate than 10 for these parameters. In the context of IDATA's focus detection problem, 5 levels is a useful resolution for both visually impaired and sighted users.

- The noise, combined roughness & noise and the visual conditions used in experiment 2 were close in terms of performance (vision being 10 % better than noise and 7 % better than combined). Therefore, for similar applications to the

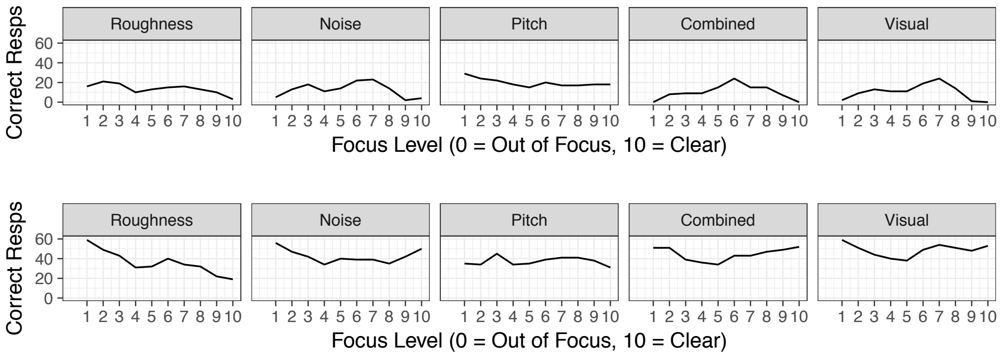


**Figure 6: Mean number of correct responses at each level, for 0 % error (top) and 10 % error (bottom).**

one presented here, these auditory parameters may provide an effective substitute for vision. This would be beneficial in contexts where the users are visually impaired - such as IDATA, or where the visual modality is already occupied by some other task.

## 6.7 Future Work

These experiments focused on a particular domain-specific problem, however there are many potential applications in which the novel acoustic parameters investigated in this paper may be useful. Any future sonification system which implements psychoacoustic sound parameters will require research to assess how these parameters can be designed to most effectively utilise their benefits.

Due to the inherent cross-modal perceptions of psychoacoustic sensations such as roughness, there is potential for research into the relationship between the acoustic parameters studied here and the cutaneous sense. It has been shown that roughness can be conveyed cutaneously through a vibrotactile device [4], so a pairing of auditory and tactile parameters may increase the performance of these parameters, allowing data sets of a finer granularity to be displayed. The combined condition in experiment 2 performed well within the 10 % error threshold - reinforcing the potential of increased performance with combining psychoacoustic parameters. Furthermore, the same audio signals used to present the stimuli aurally can be amplified and used in a vibrotactile actuator.

Now that an initial evaluation of psychoacoustic sound parameters for sonification has been carried out, investigating these acoustic mappings with visually impaired participants would be very useful knowledge to influence the tools built by IDATA. As they are aiming to involve sighted and visually impaired students, it is important to understand the perceptual responses of both user groups. Furthermore, for sonification design more generally, it is important to understand both sighted and visually impaired listeners' perceptions of the acoustic parameters used, as there can be significant deviations between the two groups [28].

## 6.8 Conclusions

In this paper, we conducted an evaluation of psychoacoustic sound parameters for a domain-specific sonification problem. We presented two studies which used various psychoacoustic parameters to display the level of focus of an astronomical image. Our first study evaluated the use of roughness (broadband noise), sharpness, as well as a combination of these two as acoustic parameters alongside the pitch mapping used by Walker [26]. Our second study evaluated roughness (pure tone), noise, a combination of the two and pitch (using a C-Major scale) as acoustic parameters, as well as a visual control condition.

We found that for all of the auditory parameters used in these experiments, there were more than twice the number of responses that were correct to within one level of focus (10 % error) than to the exact level of focus being displayed (0 % error). Therefore, this suggests that for the psychoacoustic sound parameters presented here, using a resolution of 10 levels in an auditory display is too large. A resolution of 5 levels may be provide better results.

We also found that the noise and combined roughness & noise conditions used in Experiment 2 were close in terms of performance to the visual condition. This is a promising result for the domain which is the focus of this experiment, as one of IDATA's primary goals is creating accessible astronomy data analysis tools that are as equal as possible in accuracy and functionality for both visually impaired and sighted students.

## ACKNOWLEDGMENTS

This research was funded by the EPSRC. We would like to thank Kate Meredith at Yerkes Observatory for her input from an astronomy perspective.

## REFERENCES

[1] Penny Bergman, Anders Sköld, Daniel Västfjäll, and Niklas Fransson. 2009. Perceptual and emotional categorization of sound. *The Journal of the Acoustical Society of America* 126, 6 (2009), 3156–3167.

[2] Gary Bradski and Adrian Kaehler. 2008. *Learning OpenCV: Computer vision with the OpenCV library.* " O'Reilly Media, Inc.".

[3] Stephen Anthony Brewster. 1994. *Providing a structured method for integrating non-speech audio into human-computer interfaces.* Ph.D. Dissertation. University of York York, UK.

[4] Lorna M Brown, Stephen A Brewster, and Helen C Purchase. 2005. A first investigation into the effectiveness of tactons. In *Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2005. World Haptics 2005. First Joint.* IEEE, 167–176.

[5] A L Diaz-merced et al. 2008. Sonification for the Analysis of Plasma Bubbles at 21 MHz. *Sun and Geosphere* 3, 1 (2008), 42–54.

[6] Gaël Dubus and Roberto Bresin. 2013. A systematic review of mapping strategies for the sonification of physical quantities. *PloS one* 8, 12 (2013), e82491.

[7] AA Ekdale. 2002. Sonification of paleontologic data: hearing fossil shapes. In *Abstracts with Programs, Geological Society of America*, Vol. 33.

[8] Hugo Fastl. 1982. Fluctuation strength and temporal masking patterns of amplitude-modulated broadband noise. *Hearing Research* 8, 1 (1982), 59–69.

[9] Sam Ferguson, Densil Cabrera, Kirsty Beilharz, and Hong-Jun Song. 2006. Using psychoacoustical models for information sonification. In *ICAD 2006, Proceedings of the International Conference on Auditory Display*. Georgia Institute of Technology.

[10] James W Grau and Deborah K Nelson. 1988. The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General* 117, 4 (1988), 347.

[11] S Hatano and T Hashimoto. 2000. Booming index as a measure for evaluating booming sensation. In *Proc. Inter-Noise*.

[12] Thomas Hermann, Andy Hunt, and John G Neuhoff. 2011. *The sonification handbook*. Logos Verlag Berlin.

[13] Phil N Johnson-Laird, Olivia E Kang, and Yuan Chang Leong. 2012. On musical dissonance. *Music Perception: An Interdisciplinary Journal* 30, 1 (2012), 19–35.

[14] Johan Kildal and Stephen A Brewster. 2006. Non-visual overviews of complex data sets. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*. ACM, 947–952.

[15] G Kramer, BN Walker, T Bonebright, P Cook, J Flowers, N Miner, J Neuhoff, R Bargar, S Barrass, J Berger, et al. 1999. The sonification report: Status of the field and research agenda. report prepared for the national science foundation by members of the international community for auditory display. *International Community for Auditory Display (ICAD), Santa Fe, NM* (1999).

[16] Richard Lyon. 2000. *Designing for product sound quality*. CRC Press.

[17] Sophie Maluski, Claire Churchill, and Trevor J Cox. 2004. Sound quality testing and labelling of domestic appliances in the UK. In *Proceedings of Internoise*.

[18] John G Neuhoff and Laurie M Heller. 2005. One small step: Sound sources and events as the basis for auditory graphs. Georgia Institute of Technology.

[19] Harry Ferdinand Olson. 1967. *Music, physics and engineering*. Vol. 1769. Courier Corporation.

[20] JMS Pearce. 2005. Selected observations on amusia. *European neurology* 54, 3 (2005), 145–148.

[21] S Camille Peres and David M Lane. 2005. Auditory graphs: The effects of redundant dimensions and divided attention. In *ICAD 2005, Proceedings of the International Conference on Auditory Display*. Georgia Institute of Technology.

[22] Hardial S Sagoo. 1988. *Reduction of noise levels in vaccum cleaners*. Ph.D. Dissertation. Aston University.

[23] Shin-ichi Sato, Jin You, and Jin Yong Jeon. 2007. Sound quality characteristics of refrigerator noise in real living environments with relation to psychoacoustical and autocorrelation function parameters. *The Journal of the Acoustical Society of America* 122, 1 (2007), 314–325.

[24] Jan Schnupp, Israel Nelken, and Andrew King. 2011. *Auditory neuroscience: Making sense of sound*. MIT press.

[25] A Väljamäe, T Steffert, S Holland, X Marimon, R Benitez, S Mealla, A Oliveira, and S Jordà. 2013. A review of real-time EEG sonification research. In *ICAD 2013-Proceedings of the International Conference on Auditory Display*. Georgia Institute of Technology.

[26] Bruce N. Walker. 2002. Magnitude Estimation of Conceptual Data Dimensions for Use in Sonification. *Journal of Experimental Psychology: Applied* 8, 4 (2002), 211–221.

[27] Bruce N Walker and Gregory Kramer. 2005. Mappings and metaphors in auditory displays: An experimental assessment. *ACM Transactions on Applied Perception (TAP)* 2, 4 (2005), 407–412.

[28] Bruce N Walker, Gregory Kramer, and David M Lane. 2000. Psychophysical scaling of sonification mappings. In *ICAD 2000, Proceedings of the International Conference on Auditory Display*. Georgia Institute of Technology.

[29] Eberhard Zwicker. 1961. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *The Journal of the Acoustical Society of America* 33, 2 (1961), 248–248.

[30] Eberhard Zwicker and Hugo Fastl. 2013. *Psychoacoustics: Facts and models*. Vol. 22. Springer Science & Business Media.