The Influence of Frequency-Range on the Subject's Sensitivity to Everyday Unpleasant Sounds

Max De Castro Magalhaes

Department of Structural Engineering, Federal University of Minas Gerais, Belo Horizonte, 30160 030, Brazil

DOI: https://doi.org/10.47772/IJRISS.2026.101000786

Received: 28 October 2026; Accepted: 02 November 2026; Published: 18 November 2026

ABSTRACT

The aim of this paper is to assess the influence of frequency-range on the subject's sensitivity to everyday unpleasant sounds in order to relate the processes that make them 'horrible' to their key physical characteristics (spectrum). In this paper, scraping sounds were considered on the psychoacoustics tests. The method of pair comparisons was used to measure subjects' preference orderings of 2 different stimuli, i.e. blackboard scraping and balloon noise. They were presented as discrete binary choices and in a random order to the subjects. Each one of the sounds were filtered using low and high pass filters, creating a total of eight stimuli for each everyday sound.

Keywords: Unpleasant sounds, Subject's Sensitivity, spectrum

INTRODUCTION

There are many different sounds that people find unpleasant to listen to. The archetypal worst sound that people often mention is the sound of fingernails or chalk being scraped down a blackboard or chalkboard. In the context of this research, an 'unpleasant' sound is ones that promotes an aversive reaction in a listener, and the particular interest is ones generated by a scraping action, whether that is caused by a train running badly across a set of points or the scraping of a knife on a plate.

While sounds which cause annoyance, such as traffic noise, have received considerable attention, other aversive sounds have been subject to less examination. People can have strong reactions to these sounds, for instance taking evasive action by covering their ears, and yet the reasons for these strong reactions are not understood.

One of the few studies into such horrible sounds was carried out by Halpern, Blake and Hillenbrand[1]. They examined people's responses to various horrible sounds. In the first experiment they asked 24 listeners to say how unpleasant a variety of sounds were and found that a garden tool scraped across a piece of slate was the worst. This sound is similar to finger nails being scraped down a blackboard. They then altered the frequency response of the sounds by filtering and showed that the unpleasantness came from the middle frequencies.

In a later publication, Blake[2] compared the waveform of the scraping noise with those of the warning cries of monkeys and found them to be similar. Unfortunately, it is not known what features of the waveform were compared. Blake suggested that the response to the scraping sound might be some vestigial reflex from our ancestors. The theory is that humans respond to this sound because a reflex response to monkey warning cries is still present in our brains.

In more recent research, McDermott and Hauser[3] investigated the preference of one sound over another and included a scraping sound comparable to fingernails scraping down a blackboard. They investigated cotton-top 'tamarins' and humans. The 'tamarins' reacted the same way to a screeching sound (comparable to fingernails scraping down a blackboard) as they did to amplitude-matched white noise. In contrast, humans showed a clear preference for the white noise; they disliked the screeching sound. So if the dislike of scraping sounds comes from some vestigial reflex, this reflex doesn't seem to be present in cotton-top 'tamarins'. 'Tamarins' are new world monkeys and these followed a different evolutionary path from old world monkeys, apes and humans

from about 40 million years ago. Consequently, it cannot be said that this cotton-top tamarin result disproves the vestigial reflex, because, for example, this reflex might have only developed in the last 40 million years, but it might be said to make it less likely that the vestigial reflex theory is true.

Given that the people react very strongly to scraping sounds, and the reason for this reaction is unknown, it was decided to examine the response to scraping sounds in a web experiment[4,5]. Although the experimental conditions were quite uncontrolled, using the internet allowed access to a vast number of ratings (1.5 million to date) and to a range of subjects that is virtually impossible to replicate in the laboratory. It was hoped that the web experiment on unpleasant sounds might allow some insight into whether the response to scraping sounds is a learnt reaction in some cultures or is an intrinsic part of being human. For one of the four scraping sounds tested, the variation in 'horribleness' rating with gender, age and location supported the hypothesis previously suggested by others that that the response is caused by some vestigial reflex related to the warning cries of monkeys.

However, for the actual recording of the fingernails scraping down a blackboard the results were inconclusive: the response with gender and age supporting the vestigial reflex concept, whereas the variation with location did not. An alternative hypothesis that the response was related to an audio-haptic interaction was also examined, with one test supporting the hypothesis, and the other not. Two new hypotheses for the response to this sound were forwarded drawing on work on dissonance; further work is needed to examine these.

Finally, experiments have been carried out measuring ERPs for a range of aversive sounds[6]. These results indicate an early and strong reaction to the sounds, similar to that found for emotive pictures. However, further experimentation is needed to focus particularly on scraping sounds.

THE METHOD OF PAIR COMPARISON – A BRIEF OVERVIEW

The method of pair comparison[7-9] is one of the methods known as psychological-scaling methods. The results are presented on psychological scales. All possible pairs of stimuli are considered on the test. In general, the stimuli are of similar nature and are judged according to their unpleasantness. The numerical results are given in terms of a proportion matrix which represents the proportion of the time each stimulus (at the top of the proportion matrix, i.e. on each column) is judged greater than every other stimulus at the side of the matrix, i.e. on each row.

It is simply based on a binary choice. With the method of pair comparisons a set of stimuli is judged by presenting all possible pairs of the stimuli to each subject who chooses for each pair the stimulus that better satisfies the specified choice criterion (in this case, the more unpleasant sound). One of the main advantages of the method of pair comparisons is not only its simplicity but also its use of comparative judgments. The method of pair comparisons thus utilizes our inherent familiarity with some kind of stimulus and also our ability to make comparisons. Among the methods available for eliciting comparative judgments, binary choice is considered to be the easiest one for subjects, as it compares only two stimuli at a time.

The simplest approach, which is considered herein, is to present all possible pairs of the items to each respondent. With n items, there are n (n – 1) pairs in total. Each pair results in a binary choice. Therefore, all possible combinations were considered on the experimental test. For each subject, a full set of choices yields a *preference score* for each stimulus, which is the number of times a particular noise was preferred by the subject to other stimulus considered in the set. These scores were calculated by generating a matrix (n by n), and entering a number 1 in each matrix cell where the column of particular stimulus was preferred to the row, and a 0 otherwise. Thus, for each judgment, the count matrix from each subject is summed in order to form the count matrix.

The next step is to calculate the proportion matrix. It is obtained by dividing the values on the count matrix by the number of subjects. Then, the matrix is sorted by column sum in increasing order of rank of the final scale values from left to right. In addition, the matrix is sorted by row sum, increasing from bottom to top, producing the proportion matrix.

THE LAW OF COMPARATIVE JUDGMENT

A comparative judgment is based on a choice between two stimuli. However, if they are close together, there is some confusion on choosing them. According to the law of comparative judgment, a given response separation

on a particular scale is a function of the normal deviate (or standard-measure distance between two stimuli ***j*** and ***k***), the dispersion of the two stimuli and the coefficient of correlation between them. The law of comparative judgment, also known as Thurstone's law of comparative judgment is given by [7]

$$R_j - R_k = z_{jk}\sqrt{\sigma_j^2 + \sigma_k^2 - 2r_{jk}\sigma_j\sigma_k} \quad (1)$$

where R_j and R_k are the responses to stimuli S_j and S_k respectively; $z_{jk}$is the standard-measure distance from the mean of a unit normal distribution; σ_j^2 and σ_k^2 are the standard deviations and r_{jk} is the coefficient of correlation between the responses.

The parameter z_{jk} is also known as the matrix of scale separation values. It is obtained by taking the values of the proportion matrix and fitting a curve of normal distribution to them. The remaining parameters can not be obtained directly. According to the Thurstone, some approximations can be made with respect to the law described above. Although there are five 'cases' or approximations described on ref. [7], It was assumed case III, i.e. that $r_{jk} = 0$. It means that no correlation between responses to any pair of stimuli is considered. Thus, in this case, eqn. (1) becomes:

$$R_j - R_k = z_{jk}\sqrt{\sigma_j^2 + \sigma_k^2} \quad (2)$$

where $\sigma_j = \frac{c}{V_{zj}}$ (3)

and $V_{zj} = \frac{var(Z_{jk})\,(N-1)}{N}$ (4)

$$c = \frac{N}{\sum 1/V_{zj}} \quad (5)$$

N is the total number of stimuli, *var* means variance, σ_j is the standard-deviation for the dispersion of stimulus *j*; σ_k is equal to the transpose of σ_j; z_{jk} is the scale separation matrix which is created by fitting a normal distribution curve to the values in the proportion matrix.

Thus, a new matrix can be obtained from eqn. (2). Finally, the vector containing the rating scale values R_j is obtained by summing the values on each column and taking the mean.

THE STANDARD ERROR CALCULATION VIA MONTE CARLO

The Monte Carlo procedure was used to calculate the standard error of the rating scale values for the method of pair comparison, this procedure is described by Montag and Carlson in ref. [10]. The standard error for the count matrix is first calculated over a large number of repetitions, e.g. 10000. Then, a value in the count matrix is altered by the count matrix standard error. As a result, the rating scale is modified. Next, the standard deviation of the rating scale is obtained considering all repetitions and the standard error calculated. Confidence limits are calculated for 95% confidence level.

EXPERIMENTAL TESTS

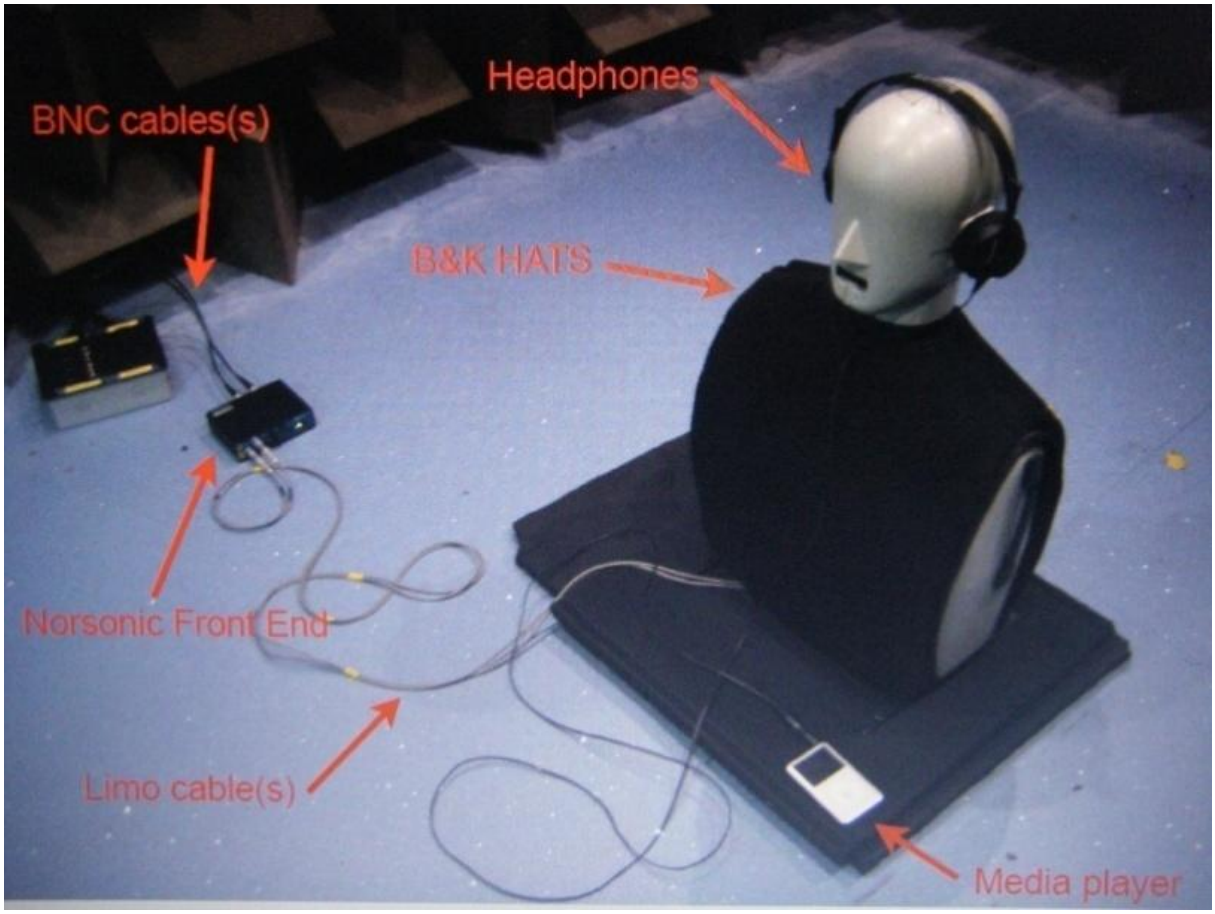
The aim of this initial work was to assess the influence of frequency range on the subject's sensitivity to unpleasant sounds. The two sounds were recorded using a Sony data recorder and an Art stereo microphone. The sampling frequency rate used was 44,100 Hz. During the recording, the input gain was set for a level range

between -6dB and -3dB. The sound files on the DAT tape were converted to WAV audio format files using a DAT player, a Yamaha digital mixing desk and a computer running Adobe Audition. Although the duration of the two recording sounds (between 3 and 5 sec.) could have been truncated in order to have the same 'length' while presented to the subjects, their duration was kept according to the original recording. It is believed that if this variable had been removed, the two sounds would have 'lost' their originality as 'everyday sounds'. The original sound with no filtering applied has frequency content ranging from 15 Hz – 16 kHz.

For the first experiment, a total of 20 subjects participated in the subjective measurements. The playback equipment used for the subjective measurements was a portable media player (iPod) and a set of headphones (Sennheiser HD 25SP), which offered a reasonable level of noise isolation. This allowed the measurements to be conducted anywhere with sufficiently low background noise levels. In the second experiment, 17 different subjects participated in the subjective measurements.

A dummy head fitted with microphones in each ear canal was used in order to measure the output level and frequency response of the playback equipment. An audio file contained a pink noise was played through the system (see Fig. 1).

Figure 1. Measurement set-up for recording the output level and frequency response of the playback equipment**.**



The playback equipment used in the second set of subjective measurements was a portable Creative Sound Blaster Card (connected to a Laptop) and a set of headphones (AKG). Similarly it allowed the measurements to be conducted anywhere. In both experiments stimuli pairs were presented in a random order.

The procedure used involves listening to sounds and rating them based on your opinion. The subjects were asked to give an opinion on the sound their hear on an scale of one to ten based on how horrible they think the sound is: one being neutral and 10 on the most horrible sound. The same rating could be used more than once. They hear three test sounds to begin with. This is to give one idea of what the test is like. Nextt, the subject will hear 8 sounds. There will be a 5 seconds gap between each sound.

In this research, the method of pair comparisons is used to measure subjects' preference orderings of 2 different stimuli, i.e. blackboard scraping and balloon rubbing. They were presented as discrete binary choices. Each of the "everyday" sounds was filtered using Butterworth type filters; three Low-pass and four High-pass filters (see Fig. 1 for the cut-off frequencies). This gave eight stimuli from each of the "everyday sounds". Subjects were instructed to indicate whether the first or second stimulus of each pair was more unpleasant. The durations of the interval between the two stimuli in a stimuli pair was 1 second. The duration of the response interval was determined by the subject's response latency. Signal generation and presentation, as well as all other aspects of data collection, were under real-time computer control.

A computer program (developed on MATLAB) was implemented for scaling pair choice data. The scaling procedure adopted was based on the law of comparative judgment[7], III case. Individual's inconsistent choices were detected using the Doubled Sorted (DS) Procedure[8,9].

As a result, those choices were re-tested, i.e. the corresponding stimulus pairs were presented to the particular

subjected again, in the end of the test. The preference scores were obtained by the summing of the columns. The matrix was then sorted by column sum, increasing from left to right. Next, the *stimuli* were arranged in an ascending order for the preference scores. Then, the matrix was sorted by row sum, increasing from bottom to top. As a result, a *double-sorted* matrix was obtained.

If this DS (double-sort matrix) contains a "1" in all cells above its principal diagonal and a "0" in all cells below the principal diagonal, this is a good indication of individual's consistent choices[8]. On the other hand, if the double-sorted matrix contains a "1" in a few cells below the principal diagonal, *circular triads* might be occurring. A *circular triad* is an intransitive preference order among three specific stimuli. For instance, three binary choices among items A, B, and C may produce the following order: $A > B > C > A$. Thus, it means that a lack of reliability can be identified and possibly minimized by re-presenting each pair of noise (corresponding to theses cells) to the subject in a random way.

The full set of pairs was presented to the subjects. It was possible once the survey was administered by computer. This allowed the inconsistent pairs to be presented again to the volunteer during the same survey ('re-test'). Thus, it might be possible to do further analysis, i.e. to determine whether the detected inconsistent responses were just a mistake or were indeed the choices the respondent intended to make. Finally, the scaling procedure adopted was based on the law of comparative judgment, case III. The rating scale values were obtained by summing the values of each column of the matrix defined by eqn. (2) and calculating the mean of their values.

RESULTS

Figures 2 and 3 show the spectrogram of two types of noise (balloon-rubbing and scraping noises) filtered using high-pass band filters at different cut-off frequencies.

Figure 4 below shows a comparison between the rating scale values (Rj) and standard errors for two separate sets of measurements using balloon-rubbing and blackboard-scraping type of noises. The values obtained were normalized to the original signal, i.e. with no filtering applied. It is seen that for the set 1 (20 subjects tested), the Rj values increases gradually with frequency. The higher levels of unpleasantness are evident for high-pass type of filters. On the other hand, for set 2 (17 subjects) the levels of unpleasantness tend to decrease as frequency increases, especially at higher frequency ranges. It is seen that the second set of results, reproduced using different subjects, showed some similarity to those published by Halpern *et al* in ref. [1], i.e. high frequencies (equal or greater than 4,000 Hz), differently from intuition, are not necessarily the main cause of unpleasantness.

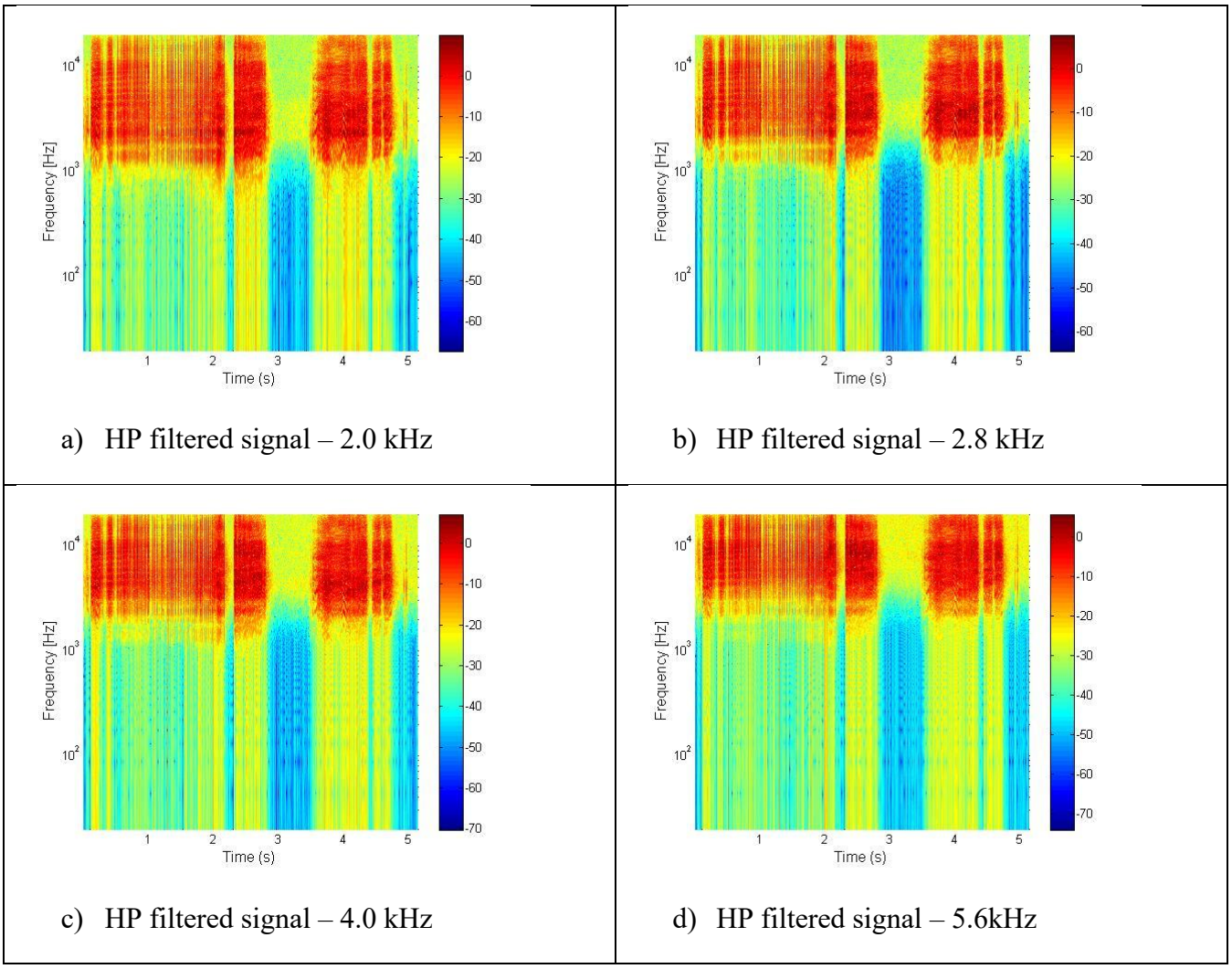
There were two main differences between experiments 1 and 2. The differences were:

i. The use of different audio equipments for the reproduction of the stimuli;

ii. Two distinct set of subjects were used for each experiment;

iii. The type of noise listened might have be identified by some of the subjects in one particular group, and it might have led to some sort of association on the subjects part;

Figure 2. Spectrogram of a balloon-rubbing noise filtered using high-pass band filters at different cut-off frequencies.



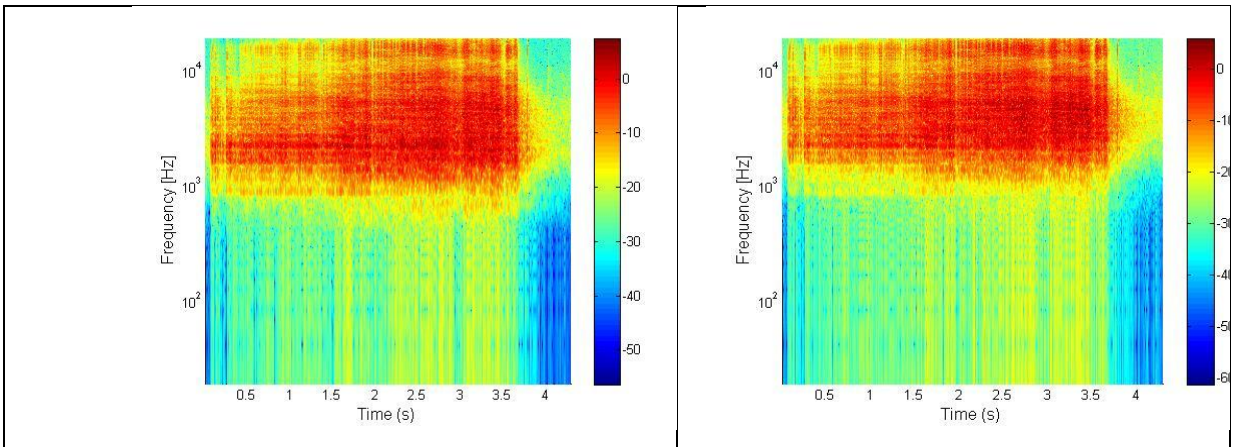
a) HP filtered signal – 2.0 kHz

b) HP filtered signal – 2.8 kHz

c) HP filtered signal – 4.0 kHz

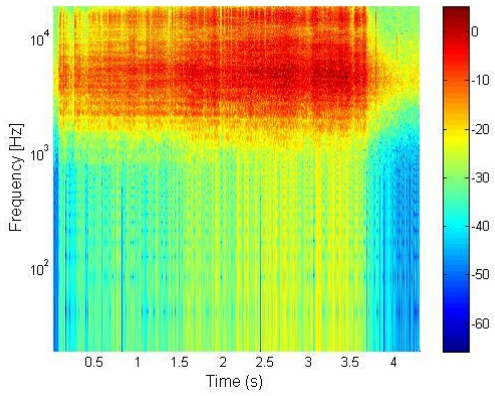
d) HP filtered signal – 5.6kHz

Figure 3. Spectrogram of blackboard noise signals filtered using high-pass band filters at different cut-off frequencies.

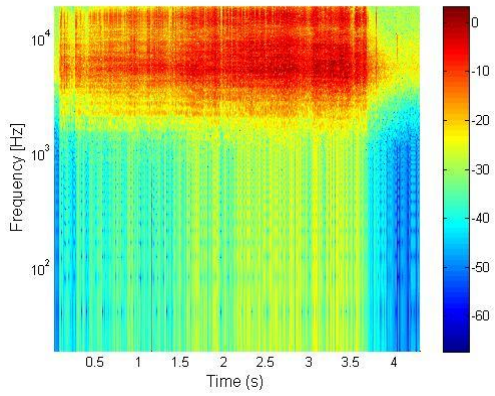


a) HP filtered signal – 2.0 kHz

b) HP filtered signal – 2.8 kHz

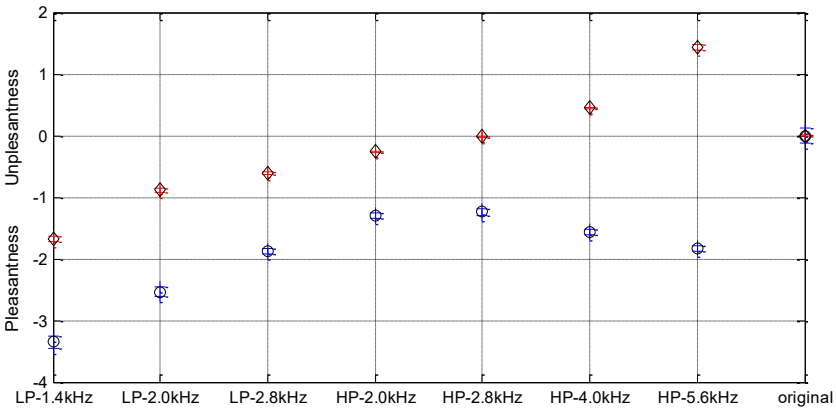


c) HP filtered signal – 4.0 kHz

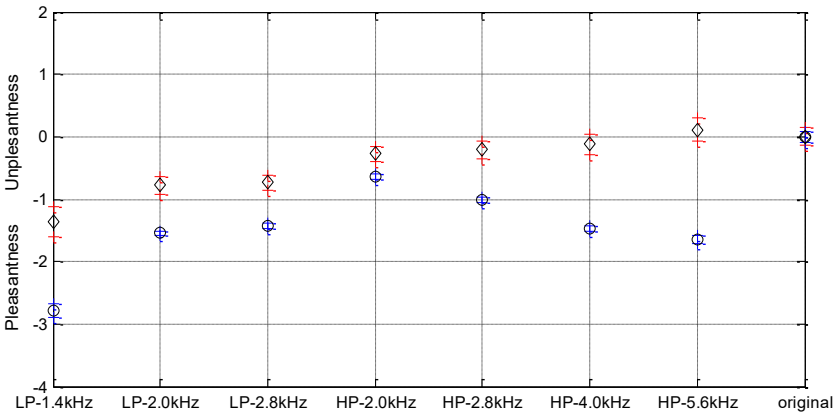


d) HP filtered signal – 5.6 kHz

Figure 4. Comparison between the rating scale values (Rj) and standard errors for two separate set of measurements filtered using 3 LP and 4 HP filters; ◊ - set 1 (20 subjects); ○ – set 2 (17 subjects); a) balloon noise; (b) blackboard noise.



a) Balloon noise



b) Blackboard noise

CONCLUSIONS

The research on unpleasant sound and how they are perceived indicates that the spectrum of a particular noise signal might be related to its perceived unpleasantness. If the common reaction to unpleasant sounds is evolutionary, it could be linked to the perception of consonance and dissonance.

The association that a subject makes, if any, and how this association affects their choices are an unknown to everyone, but the subject. Thus, the identification of a particular type of noise may lead to some ambiguity.

According to the results it is seen that for high-pass filters the two set of tests present different trends as frequency increases. Apparently, the results suggest that the main reason for the discrepancy between the set of results is the use of different equipments for the sound reproduction. Thus, it might be possible to do further analysis, i.e. to determine whether the detected inconsistent responses were just a mistake or were indeed the choices the respondent intended to make.

The results might not just be useful because they can explain the human condition better, but also because they would enable noise control regimes to be designed to attenuate scraping sounds from mechanical sources such as train wheels on tracks.

A larger sample of subjects may help to understand the uncorrelated results at higher frequencies.

ACKNOWLEDGEMENTS

This research was funded by CAPES (Research Foundation subordinated to the Ministry of Education of Brazil) and University of Salford (UK).

REFERENCES

1. Blake, R. (1986). Margins - primal screech. *Psychology Today*. **20** (9). 68.
2. Brown, Thomas C.; Peterson, George L. (2007). An Enquiry into the Method of Paired Comparison: Reliability, Scaling, and Thurstone's Law of Comparative Judgment. *Rocky Mountain Research Station, U.S. Forest Service Report*: RMRS-RWU-4851.
3. Brodersen, R. Marc; Ford, Valerie; Bell, Paul A. (2005). The judged seriousness of an environmental loss is a matter of what caused it. *Journal of Environmental Psychology* **25** (1): 13-21.
4. Czigler I, Cox TJ, Gyimesi K, Horvath J, (2007). Event-related potential study to aversive auditory stimuli. *Neuroscience Letters* 420 (3), 251-256.
5. Guilford, J. P. (1954). *Psychometric methods* New York. McGraw-Hill.
6. Halpern, D. L., Blake, R. and Hillenbrand, J. (1986). Psychoacoustics of a chilling sound. *Percept. Psychophys*, 39(2), 77-80.
7. McDermott, J. and Hauser, M. (2004). Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. *Cognition*, 94 (2). B11-B21.
8. Montag E. and Carlson C. (2006). Empirical formula for creating error bars for the method of paired comparison. *J. Elec. Imag.*, **15** (1) pp 10502, 1-3.
9. T. J. Cox (2007). Bad vibes: an investigation into the worst sounds in the world. *PPA-09-003, proc. 19th ICA* Madrid.
10. T. J. Cox (2008). Scraping Sounds and Disgusting Noises. *Applied Acoustics*. v. 69 (12), 1195-1204
11. Fusaro,G., Kang J., Asdrubali, F. Chang, W.S. (2022). Assessment of acoustic metawindow unit through psychoacoustic analysis and human perception. *Applied Acoustics*. v. 196, 1-11.
12. Effect of source azimuth on the perceived annoyance of domestic sounds (2027). *Applied Acoustics*. v. 257 (12), 1195-1204