

# Scraping sounds and disgusting noises

Trevor J. Cox \*

*Acoustics Research Centre, University of Salford, Salford M5 4WT, UK*

Received 22 May 2007; received in revised form 8 November 2007; accepted 12 November 2007

Available online 8 January 2008

## Abstract

Thirty-four horrible sounds have been examined in an Internet-based psychoacoustic experiment. This paper presents the results for the scraping and disgusting noises used. It is not understood why some humans find certain scraping noises, such as the sound of fingernails being scraped down a blackboard, so terrible. In this experiment, the variations in ratings with age, gender and location are examined. The results for one of the scraping sounds is consistent with the hypothesis suggested by others, that the response comes from a vestigial reflex related to the warning cries of monkeys. But this was not true for the actual recording of the fingernails scraping down a blackboard. An alternative hypothesis that the response is related to an audio–haptic interaction was tested and results indicated that this idea warrants further investigation. Other possible causes of the response, drawing on work concerning dissonance, are tentatively suggested. The disgusting sounds examined included the worst sound found in the experiment, the sound of someone vomiting. However, none of the disgusting sounds tested promoted responses consistent with a ‘disgust reaction’ based purely on survival instincts. Cultural factors might be important in our response to the disgusting sounds, with the influence of manners and etiquette being suggested as a possible factor.

© 2007 Elsevier Ltd. All rights reserved.

*Keywords:* Aversive sounds; Noise; Horrible sounds; Web experiment; Fingernails; Blackboard; Disgust

## 1. Introduction

There are many different sounds that people find horrible 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. But there are many other sounds, such as the sound of someone vomiting, which are also horrible. In the context of this paper, a horrible sound is one that promotes an aversive reaction in a listener.

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 horrible sounds, for instance taking evasive action by covering their ears, and yet the reasons for these strong reactions is not understood. The intention of the study reported in this paper was to provide

information about how gender, age and location, influence listeners ratings of a variety of horrible sounds. And from the relationships between these variables and horribleness, to try and infer what makes people find these sounds horrible. The research involved a web based experiment where people auditioned and rated horrible sounds. Most of the results presented in this paper are based on 385,000 ratings from the experiment.

A brief outline of the method is given in the next session, after which the voting trends for all horrible sounds as a group is considered. Then the response to individual sound types is considered in subsequent sections. Section 5 examines scraping sounds and tries to see if the evidence indicates an evolutionary or haptic basis to peoples’ response to sounds such as the sound of fingernails scraping down a blackboard. Section 6 examines sounds which, during the experimental design, were expected to create a disgust reaction. As the results will show, however, the response to these sounds do not follow the expected pattern for a disgust reaction purely based on a survival instinct.

\* Tel./fax: +44 161 295 5474.

E-mail address: [t.j.cox@salford.ac.uk](mailto:t.j.cox@salford.ac.uk)

## 2. Method

Testing human responses to stimuli via the Internet has become increasingly popular in many areas of psychology. While a methodology which uses the Internet is fraught with difficulties, the opportunity to test large data sets across a wide range of subjects is highly appealing. Nowadays, nearly all computers are sold with the ability to reproduce sound and to connect to the Internet and the number of broadband connections is rapidly increasing. Consequently, it is suggested that it is now possible to carry out psychoacoustic tests across the Internet where people actually audition and judge sound files [1].

The experiment described here was aimed at the general public. Consequently, the method had to be simple, unambiguous and appealing. From a user's perspective, the experiment was as follows. When users first went to the website, they were asked for a few details about themselves: their gender, age (within 10-year age ranges) and location. This was to give data to be able to interpret the voting patterns. A cookie was used to store this information on the subject's computer.

In future experiments, it would be good to also ask for some simple description of the sound reproduction system being used, e.g. laptop loudspeakers, headphones or computer loudspeakers, because this might be a significant contextual variable which would be easy to obtain.

Next participants were presented with the "sound-check" screen to ensure the sound on the computer was turned on and that the reproduction level was reasonable. A sample of speech was presented which said: "set the volume level so you can hear me speaking clearly, as though I was having a conversation with you". One of the problems with carrying out psychoacoustic experiments on the web, is the lack of control over the loudness of the sound reproduced. In most perceptual experiments, the loudness of the sound will have a significant effect on people's judgements. This sound-check screen was intended to help reduce the variation in the volume levels. Even with this precaution, however, no proper calibration for the reproduction level could be achieved and it should be assumed that subjects listened to the sounds with a variety of volume levels. It is assumed that this causes a significant additional error, but if this is a random error, and the effects of level are simple (e.g. linear), then by getting sufficient numbers of subjects, it should be possible to look for underlying trends for an average listening level. However, if there is an interaction between listening level and other factors, say age, then this could also introduce a bias into the results. However, this bias can be minimised by using appropriate analysis. Overall, the lack of calibration is a significant drawback to carrying out psychoacoustic experiments on the web. Tests need to be carried out to compare results from the web to laboratory experiments to test whether these assumptions about the 'averaging' out of error is correct.

Next, participants came to the voting screen. Users pressed play, listened to the sound and then voted using

a direct scaling method on an ordinal scale. The use of a relatively short ordinal scale has consequences on the analysis technique. The voting patterns were distinctly non-normal, and therefore non-parametric analysis methods have had to be used. A box in the middle of the screen either displayed an image or was blank. The results from the votes were stored in a mySQL database. The IP address of the computer was also stored, and this allowed us to estimate how many votes had been cast by each user. Sometimes the number of votes were very large, and is likely to be indicative of many people using the same machine.

People were given the chance to vote at anytime while listening to the sound file. It is suspected that sounds which were horrible from the beginning of the recording ranked higher. In future experiments, a minimum listening time will be introduced before votes can be cast to reduce rating variance introduced by length of listening time.

Using the web to run experiments offers a number of advantages, but also a series of methodological challenges [2,3]. As noted before there is no calibration of levels and the quality of the reproduction equipment varies between subjects. Subjects are self-selecting and the context of listening is uncontrolled. However, in a sense self-selection and listening context are no more artificial than is achieved in many laboratory experiments. As scientists, we are used to working in laboratories, but to subjects in perceptual experiments, listening rooms and anechoic chambers are strange artificial spaces.

After the contextual data was gathered, the listeners auditioned a series of horrible sounds presented in random order which they then graded on a six point ordinal scale from not horrible to horrible. There were 34 sounds used in the experiment. To minimise download times, the sound files were mono and short, and they were compressed with MP3 coding at 96 kps. The sounds were set up to loop indefinitely.

## 3. Voting trends across all sounds

A previous paper gave some summary statistics for the contextual variables for the web experiment [1]. A Kruskal-Wallis test [4] was used to rank order the sounds over 487,335 votes. Table 1 summarises the rank ordering. A paired-comparison method was used to group the sounds. Although the rank order is interesting, the ordering not only reflects how horrible the sound is, but also other factors, such as how good the sound recording was. For instance, the vomiting sound was a particularly good (or awful) rendition. Consequently, the following analysis concentrates on relative changes with age, gender and location.

A generalized linear model for ordinal data was used in analysis. There are a number of contextual variables; sound code number  $s$  ( $1 \leq s \leq 34$ ), age category  $a$  ( $1 \leq a \leq 7$ ), gender  $g$  ( $g = 0 \vee 1$ ), location  $L$  ( $1 \leq L \leq 8$ ), and number of previous auditions by the respondent,  $n$ , for which we are trying to explain the variation in a horribleness rating,  $H$ . A model is fitted that finds the variation in horribleness

Table 1  
The sounds used in the experiment rank ordered

Rank	Sound	Rank	Sound	Rank	Sound	Rank	Sound
1	34. Vomiting	9=	13. Mains hum	16=	21. Polystyrene	28=	8. Electrical throb
2	32. Microphone feedback	11	25. Tasmanian devil	20	11. Dentists' drill	28=	14. Cat eating noisily
3=	19. Multiple babies	12=	29. Cough	21	10. Cough & spit	30	2. Whoopee cushion reverberated
3=	3. Scrape/squeak (like train wheels)	12=	18. Cat spitting and howling	22	23. Alarm clock	31	26. Aircraft take-off
5	1. Squeak (sounds like a seesaw)	12=	31. Mobile phone rings	23	12. Fast electrical drilling	32	24. Drums
6	22. Violin	15	16. Creaky door	24	17. Apple munch	33	6. Gong
7=	4. Whoopee cushion	16=	30. Barking mad dog	25	15. Creaky door	34	5. Low not-quite-eerie noise
7=	7. Baby cry	16=	27. Sniff	26=	9. Squeaky trolley		
9=	28. Soap opera argument	16=	20. Fingernails scraping down a blackboard	26=	33. Snoring		

rating due to the contextual variables. The following proportional odds model was [5]:

$$\text{logit}[P(H \leq j)] = \alpha_j + \lambda_s + \beta_L + \gamma_g + \delta_a + \psi n + s[\xi_a + \nu_L + \tau_g] \quad (1)$$

The logit is the log of the odds of a response in horribleness category  $j$  or below.  $\alpha_j$  is the intercept for each of the six response categories on the 'horribleness' rating scale.  $\lambda_s$ ,  $\beta_L$ ,  $\gamma_g$ ,  $\delta_a$  and,  $\psi$  are constants to explain the variation in the logit for sound number, location, gender, age category and vote number respectively.  $\xi_a$ ,  $\nu_L$  and  $\tau_g$  are constants to explain the interaction between the sound number and the age category, location and gender. A number of different link functions were tried to see which worked best for the data, and it was found that the logit function was most suitable. The values of the constants in Eq. (1) which give predicted logit values closest to those for the actual responses are found using a maximum likelihood procedure. There are 515 unique constants to be found for this model. A number of different proportional odds models were tested with a variety of constants to explain interactions within the data. The model in Eq. (1) was used because it was relatively compact and explained a good proportion of the data variation.

The  $-2$  log likelihood dropped from 274469 for the intercept-only model to 174447 for the full model showing the usefulness of the contextual variables in explaining some of the data variation. However, the pseudo- $R^2$  values indicate that the models used typically account for only 22–23% of the variation in the data. This indicates that the contextual variables available do not account for a large amount of the variation. This is unsurprising, because the experiments were conducted over the Internet, and consequently the data was produced in a not terribly controlled manner and many contextual issues, e.g. reproduction level, are ill-defined. The deviance suggested a good fit whereas the chi-squared statistic indicates a poor fit. (The test for parallel lines normally used for this type of model was not used because it is unreliable for this size of dataset, however a visual inspection indicated that the assumption

was probably reasonable.) Overall, the statistics concerning the model fit suggest the data is very 'noisy'.

#### 4. Scraping sounds

Some people have a strong reaction to certain sounds; indeed the archetypal horrible sound is the sound of fingernails scraping down a blackboard. One of the few studies into such horrible sounds was carried out by Halpern et al. [6]. 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 fingernails 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 [7] 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 strong response to scraping sounds might be some vestigial reflex from our ancestors. He suggested 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 [8] 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 does not 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

Table 2  
The rank ordering of locations from the one which have the highest ratings (left) to the smallest (right)

South America	UK	Africa	Australia	Rest of Europe	Middle East	North America	Rest of World

The shading indicates groupings of locations which are not statistically significantly different.

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 this in the web experiment. A variety of sounds were chosen which include an actual recording of fingernails being scraped down a blackboard, polystyrene being rubbed together and metal being scraped against metal. Because there was little previous work to base the selection of the sound samples on, the experiment was explorative in the sense a wide variety of sounds were chosen to be examined.

It was hoped that the web experiment on horrible 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. Curtis et al. [9] used a similar approach and arguments when examining disgust reactions based on images. Following this approach, if the response to the scraping sound is intrinsic to being human, it might be expected to operate similarly across cultures, and so the ratings should not vary with location. Furthermore, if the response follows Blake's hypothesis that it is a vestigial response concerning warning cries, then it might be expected that the response should: (a) be stronger in females because they play a role in protecting both themselves and their offspring from attack, and (b) decrease with age as an individual's reproductive potential declines.

In general, the data for age and gender is more robust than that for location. Locations were gathered as Australia, Africa, Middle East, North America, Rest of Europe, Rest of World, South America and UK. (The reason for separating the UK from the Rest of Europe was because this was a UK-based study and much of the media attention and consequently participants were expected to come from the UK.) The concept behind gathering location data is to get a sense of how responses varied with culture. However, this makes an assumption that the geographical variation in responses maps directly to the cultural values of the noises (Table 2).

Another theory concerning scraping sounds is that there could be a link between hearing the sound and how it feels to drag one's fingers down a blackboard, and this could provoke the aversive reaction. In the listener's mind, is there a link between the sound and the unpleasant touch sensa-

tion? In an attempt to gain an insight into this, audio-visual interactions were examined. A separate experiment was carried out on the website where images were shown as the sound played. As the fingernails scraping down the blackboard sound was played, one of four images were shown as illustrated in Fig. 1. In analysis, the images are considered in pairs with the responses to the two cushions being compared, and then the responses to the two metal discs being compared. The images in each pair were meant to invoke a different tactile interpretation, while maintaining other factors such as colour balance. The images were deliberately chosen to not have any obvious association with the sounds used. The hope was that the different tactile interpretations would interact with the responses to the sound. A control sound was also used, sound 25 Tasmanian devil, so respondents might also see the four images with the Tasmanian devil image. This control sound was chosen because it had a similar location in the rank ordering of the sounds as the fingernails scraping down the blackboard sound.

#### 4.1. Analysis

The analysis used makes a comparison between the scraping sound being examined and the 'average' response

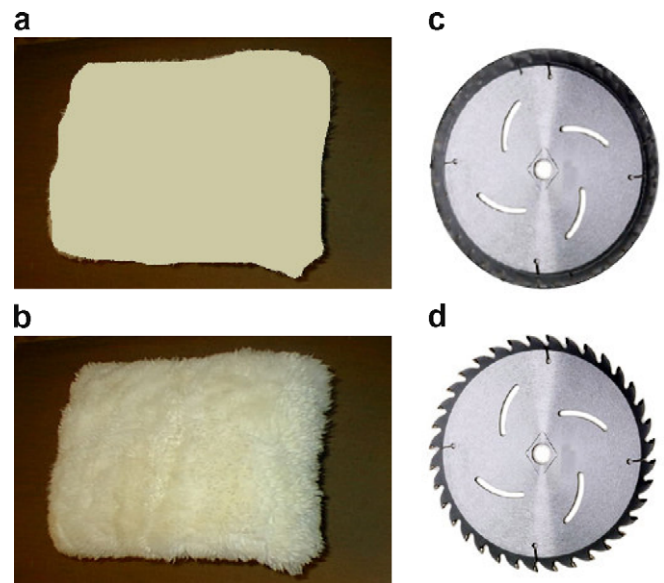


Fig. 1. The images used to examine the tactile component of fingernails scraping down the blackboard sound: (a) de-fluffed cushion; (b) fluffy cushion; (c) metal disc; and (d) saw blade.

to all the other horrible sounds. This reduces any effects due to the way different groups, such as different age groups, use the rating scales, but is reliant on an assumption that the ‘average’ response to all other horrible sounds is a meaningful comparison group. The analysis used a simplified proportional odds model where instead of fitting a different parameter for each sound,  $\lambda_s$ , a single binary parameter was used which was 1 for the scraping sound being considered, and 0 for all other sounds. This was done because it explicitly forces the comparison between the scraping sound and all other sounds and simplified the interpretation of the results from the proportional odds model.

#### 4.2. Results

Sound 20 was a recording of fingernails being scraped down a blackboard and was examined first. Interestingly, although many people cite this as the archetypal horrible sound, this recording actually only came midway in the rank listing of all sounds shown in Table 1.

Females found this sound slightly worse ( $\tau_g = 0.075 \pm 0.072$ ,  $p = 0.039$ ). Fig. 2 shows the effect of age on the response to sound 20 in comparison to the other sounds. People with ages 15–35 score sound 20 significantly worse than older and younger people (significance ranges from  $p < 0.0001$  to  $p = 0.043$  depending on which age ranges are compared) There is also a significant linear drop off with age from 35 to 65 ( $p < 0.0001$ ).

The decrease in horribleness with age for older adults raises the question as to whether the natural decrease in hearing acuity with age might cause the effect. However, Halpern et al. [6] showed that high frequencies were not the most important bandwidth in the rating of horrible scraping sounds. Consequently, it is assumed that presbycusis is not the reason for the reduction of horribleness with age. It could be argued therefore, that the decrease in horribleness with age supports Blake’s vestigial reflex hypothesis because people become less sensitive to the sound as an individual’s reproductive potential declines.

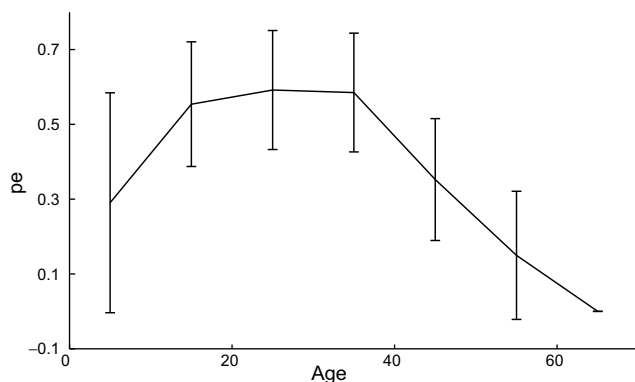


Fig. 2. The difference in logit value for scraping sound 20 and all other sounds as a function of age.

The gender result also supports this, but the significance is rather marginal.

It is possible to get some insight into the cultural importance of this sound by examining the work done on audio–visual interaction and reported in a previous paper [1]. In a separate experiment to the one reported in this paper, the sounds were presented with different visual stimuli: an associated image, an unassociated image or a blank square. For instance, the sound of fingernails being scraped down a blackboard was presented with the image of a hand on a blackboard (associated), some Smarties (unassociated) or a blank (green) square. There were two sounds where the associated image made a much bigger difference to the rating than other noises; one was the dentist drill and the associated image of a dentist, the second was the sound of fingernails scraping down the blackboard and a picture of a hand on a blackboard. In both cases the image made the sound more horrible. When the audio–visual data for the fingernails being scraped down a blackboard was examined across location, it was seen that a similar increase in horribleness response with the image is invoked in Australia, North America, Rest of Europe and UK. This implies that the mystique of this horrible sound is present in all these countries. For other locations, the amount of data was too small and the uncertainties too large to draw definite conclusions because the audio–visual interaction experiment was only run for a short time on the website.

Returning to the results without images, some variations in horribleness with location were found: Australia, Rest of Europe and South America scored sound 20 lower than the UK ( $p = 0.040$ ,  $< 0.0001$ ,  $0.011$ ), and North America scored the sound higher than the Rest of Europe ( $p = 0.0086$ ). This is more indicative of the response to the sound not being intrinsic to being human, or alternatively, evidence that the intrinsic response is modified by cultural factors as has been seen with responses to dissonance [10]. The response to sound is stronger in the UK and North America than some other parts of the world, e.g. the Rest of Europe. This stronger reaction does not appear to be a consequence of the mystique behind the sound, because in the UK, North American and Rest of Europe listeners all responded equally strongly to the image of the sound.

The examination of the results for the tactile images shown in Fig. 1 yielded a mixed response. Just over 20,000 cases for each of the two sounds examined were used, with sets of respondents chosen which were matched for the other contextual variables. A Kruskal–Wallis test showed that there was no significant variation of responses with the image presentation for the control sound, sound 25 Tasmanian devil. For sound 20, fingernails being scraped down the blackboard, there was a significant variation with image ( $p = 0.005$ ). Mann–Whitney tests showed that there was no significant difference between the fluffy and de-fluffed cushion, but there was a significant difference between the saw blade with and without teeth, with the presence of teeth making the sound less horrible. Unfortunately, given this mixed result, it is rather hard to draw any

definite conclusions, except to say it may be an avenue worthy of further investigation.

There were other sounds in the web experiment that were metal scraping sounds: 3. Scrape/squeak (like train wheels) and 1. Squeak (sounds like a seesaw) which were ranked high up the list, much higher than the actual recording of fingernails being scraped down the blackboard. Sound 9, Squeaky trolley appeared towards the bottom of the list. These other sounds were examined to see if the results were similar to, or different from sound 20, Fingernails being scraped down a blackboard. All three sounds (1, 3 and 9) were worse for females ( $p \approx 0.036$  for all sounds) as was found for fingernails scraping down the blackboard. A similar response with age was found for all these sounds with a significant linear drop off for ages 35 and above, and a raised response for ages 15–35. However, for location, the results were different when compared to sound 20. The proportional odds model showed significant effects for the Middle East, but this more reflects the fact that the number of responses from the Middle East was rather low and so the proportional odds model could not discriminate a significant difference for the Middle East. Ignoring the Middle East, for sound 3 there was no variation between other locations, for sounds 1 and 9 there were some significant variations, but not the same ones as found for sound 20. So overall, there is no clear pattern with location.

Sound 3 therefore most conforms to Blake's vestigial reflex theory, as it is worse for females, worse for ages where reproduction is more likely and similar across cultures. It also comes closest to fitting the mystique for scraping sounds to be the worst in the World by coming third in the ranking list.

Scraping polystyrene, sound 21, was found by Halpern et al. [6] to be the second worst sound in their study. In the web experiment, it came midway down the rank list as shown in Table 1. Gender was not a significant effect for this sound. Like other scraping sounds, there is a drop off with age, but starting from a younger age 15 (logit gradient =  $-0.0247 \pm 0.0034$ ,  $p < 0.0001$ ). There was little variation with location, the only significant difference being that North America rated the sound more horrible than the UK ( $p = 0.016$ ).

#### 4.3. Discussion

Gaver [11] and Dubois [12] showed that people describe everyday sounds by trying to associate the sound with a source or a meaningful event. If the source or event is identifiable, then a respondent's description of a sound is likely to be dominated by the source or event, rather than the properties of the signal. If the sound is not identifiable, then descriptions relating to the physical characteristics of the signal are important. A key question is whether the response to scraping sounds is something intrinsic to the signal, or something more to do with the identification and association of the sound to a source or event, as would

be the case in some form of haptic–acoustic interaction. The scraping sounds which gave the highest scores, sounds 1 and 3, are unlikely to be identified as fingernails scraping down a blackboard, because they clearly sound different. (No formal tests of whether they sound different were undertaken, because it seemed pointless given that the differences were self-evident even to a casual listener.) Consequently, this adds weight to the argument that there is something intrinsically unpleasant about scraping sounds, and the response is not just about association. Especially, as over recent decades blackboards have become much less common as they have been replaced by white boards. The results for event related potentials (ERPs) and horrible sounds tentatively support this [13]. The ERP results are similar to those seen for emotionally arousing pictures, with surprisingly early negativity that might be an initial effect in a broad neural network including limbic structures. However, the results from the Internet experiment failed to produce clear evidence to support Blake's vestigial reflex hypothesis. So, could there be another explanation for the strong response to these sounds?

One aspect of sound perception which has had considerable attention is dissonance and consonance of musical notes. The evidence is that musical consonance is influenced by social and cultural factors, and what is heard as consonant can be changed by learning. However, dissonance seems much more robust across cultures, and there are certain combinations of pure tones that seem intrinsically unpleasant. The exact reasons for the response are unknown, but several authors have made suggestions. One theory is that dissonance is a by-product of the auditory system being trained to understand speech, in particular a by-product of extracting the harmonic components of speech when there is background noise present [14]. Another hypothesis is that these preferences have arisen as part of music-specific adaptations because they exist only in humans and not in non-human primates [15]. A third theory is that a dissonant stimulus reduces our capacity to hear other sounds, and therefore we have evolved to find this experience unpleasant because this leads to stimulus-aversive behaviours [16].

These hypotheses as to why humans find certain sound dissonant could equally be applied to why scraping sounds are unpleasant. Consequently, it is suggested that there are two further possibilities for why scraping sounds are horrible: the response may have arisen as a by-product of our language learning, on the other hand it might be an evolved response to avoid stimuli which are aversive to hearing other sounds. The Internet experiment reported in this paper cannot provide any evidence as to the validity or either of these suggestions, and examining these possibilities is left to future work.

#### 5. Disgusting sounds

Disgust is a common reaction to many things such as bodily excretions and secretions [17,18]. Humans find bod-

ily excretions and secretions disgusting because they might contain high concentrations of pathogens. There is good evidence that humans evolved to have a disgust response to avoid disease and illness. Disgust can lower blood pressure, cause nausea, and make people take evasive action. Many disgusting sights, such as someone spitting, are also associated with unpleasant sounds. So this part of the experiment aimed to examine whether certain horrible noises provoke a ‘disgust reaction’. Consequently, the eight sounds examined relate to bodily functions which might be associated with bodily excretions and secretions: 34 Vomiting, 27 Sniffing, 17 Eating an apple, 14 Cat eating, 2 Whoopee cushion, 4 Reverberated whoopee cushion, 29 Coughing and 10 Coughing and spitting. A literature survey indicated that disgust and sounds have not been investigated before.

As stated previously, Curtis et al. [9] found that responses to disgusting pictures was stronger in females, probably because females play a double role in protecting both self and offspring from disease. They also found that the disgust response decreased as people got older and was similar across cultures. If the rating of the horrible sounds is dominated by a disgust reaction related to disease avoidance and survival, it might be expected that similar trends would be found in the web experiment described here because it follows a very similar methodology.

However, there is also a strong social element to disgust; for instance, someone might be disgusted by immoral or unfair acts [19]. It is suggested that there may be certain confounding issues around manners and etiquette which might make the responses to the sounds not purely about disgust as a means to avoid disease.

### 5.1. Results

The first analysis examined the ‘average’ responses across all eight disgusting sounds using a proportional odds model. Females found the disgusting sounds significantly more horrible than the males ( $p < 0.0001$ ). There is a linear decrease in horribleness from ages 5 to 25 (logit gradient =  $-0.011 \pm 0.007$ ,  $p = 0.0008$ ), and an increase in horribleness from ages 25 to 64 (logit gradient =  $0.0071 \pm 0.0016$ ,  $p < 0.0001$ ). The only variation in location is that South America find disgusting sounds less horrible than Australia, North America, the Rest of the World and the UK ( $p = 0.044$ ). Therefore, the hypothesis that these sounds are a set which all invoke a disgust reaction, and give ratings which behave in a similar way as measured by Curtis et al. [9] is not supported by the results. Most telling is the general increase in the horribleness of sounds as adults get older, contrasting with the visual disgust experiment which showed a decrease in disgust as reproductive potential declined.

Proportional odds models were then fitted separately for each of the disgusting sounds and the variation in the rating with gender, age and location were examined. The results varied between sounds. For instance, sound 14

Cat eating has a linear decrease in horribleness with age (logit gradient for ages 15–65:  $-0.0040 \pm 0.0033$ ,  $p = 0.007$ ), whereas sound 29 Coughing has a linear increase in horribleness with age (logit gradient ages 5–65:  $0.0157 \pm 0.0036$ ,  $p < 0.0001$ ). None of the sounds produced the anticipated response (females worse than males, decreasing horribleness with age, responses similar across locations). It appears that disgust and disease avoidance are not the only factors being considered when the listeners rated the sounds.

Table 3 summarises the results from the analysis of these sounds. The most horrible sound was 34 Vomiting as shown in Table 1. Females rate this sound more horrible than males, there was only a little variation with age, and the most significant pattern in the location results was that the UK found this much more horrible than most other countries. The sound recording was clearly the sound of someone vomiting, so the results can not be explained by misinterpretation of what the sound was. Indeed, it was surprise that this did not produce results consistent with the disgust and disease avoidance hypotheses, especially as sick is normally one of the strongest stimuli for evoking disgust.

There were two eating sounds: 17 Eating an apple and 14 Cat eating. These are sounds that might be open to different interpretations by listeners. Both were clearly eating sounds, but people might not know if it was a human or another animal eating and the reaction might depend on this. In the case of 17 Eating an apple, this is the sound that gives results most similar to the disgust hypothesis. It is worse for females, the biggest response is for ages where people are most likely to have children, but there is some variation with location as noted in Table 3. The clearest trend for 14 Cat eating sound is the linear decrease in horribleness with age.

Only midway through auditioning 27 Sniffing, does it become obvious what the sound is of, because then the actor used in the recording blew their nose. Consequently, people who rated this noise only based on the early part of the sound, may have interpreted it as being something other than sniffing, because the sound was hard to identify. Consequently, the lack of many trends for this sound may be due to additional response variations induced by uncertainty in identification.

There were two coughing sounds, 29 Coughing and 10 Coughing and Spitting and the sources of the sound in both recordings were easy to identify. The responses showed a rise in horribleness with age for both cases; opposite to the expected direction for a disgust reaction. This could show the influence of manners; maybe older people are less tolerant of public coughing. The common trend in the location results for the two sounds is like vomiting, the UK finds these more horrible. Other variations with locations are indicated in Table 3. Females found sound 10 worse than males, which was the sound of a male coughing. Sound 29 was a female coughing, and in that case both genders gave similar ratings.

Table 3  
Most of the significant trends in the sounds expected to produce a disgust reaction

Sound	Gender	Age	Location
34. Vomiting	F ( $p < 0.0001$ )	Logit 45 > 65 ( $p = 0.05$ ) Logit 55 > 35 ( $p = 0.03$ ) Logit 55 > 65 ( $p = 0.02$ )	Logit UK > Australia, Middle East, North America, Rest of Europe, Rest of the World, South America ( $0.0001 > p < 0.005$ ) Logit North America > South America ( $p = 0.04$ )
17. Eating an apple	F ( $p = 0.0001$ )	logit 35 > 15 ( $p = 0.02$ ) logit 35 > 65 ( $p = 0.01$ ) 35-65: logit gradient = $-0.0062 \pm 0.0057$ ( $p = 0.013$ )	Logit South America < Australia, North America, Rest of Europe, Rest of the World, UK ( $0.0001 > p < 0.009$ ) Logit Rest of the World < UK ( $p = 0.046$ )
14. Cat eating	–	15–65, logit gradient = $-0.0040 \pm 0.033$ ( $p = 0.007$ ) Logit 5 > 15 ( $p = 0.00015$ ) Logit 5 > 25, 35, 45, 55, 65 ( $p < 0.0001$ )  Logit 15 > 65 ( $p = 0.0001$ ) Logit 35 > 65 ( $p = 0.03$ )	Logit Australia > Middle East, North America, UK ( $p = 0.04, 0.01, 0.04$ ) Logit Middle East < North America ( $p = 0.04$ ) Logit North America > Rest of Europe, UK ( $p = 0.0003, < 0.0001$ ) Logit Rest of Europe < Rest of the World ( $p = 0.05$ )
27. Sniffing	–	–	Logit Middle East < Australia, North America, UK ( $p = 0.04, 0.04, 0.008$ ) Logit UK > North America, Rest of Europe, Rest of World, South America ( $p < 0.0001, < 0.0001, 0.005, 0.035$ ) Logit North America > Rest of Europe ( $p = 0.04$ )
10. Coughing and spitting	F ( $p < 0.0001$ )	5–65, logit gradient = $0.0132 \pm 0.0042$ ( $p < 0.0001$ )	Logit Australia > North America, Rest of Europe, South America ( $p = 0.04, 0.039, 0.006$ ) Logit UK > North America, Rest of Europe, Rest of the World, South America ( $p < 0.0001, < 0.0001, 0.04, < 0.0001$ )
29. Coughing	–	5–65, logit gradient = $0.0157 \pm 0.0036$ ( $p < 0.0001$ )	Logit UK > North America, Rest of Europe, Rest of the World, South America ( $p = 0.015, 0.002, 0.002, 0.016$ ) Logit Middle East > Rest of Europe, Rest of World, South America ( $p = 0.046, 0.014, 0.03$ ) Logit North America > Rest of World ( $p = 0.044$ )
2. Whoopee cushion	–	5–35, logit gradient = $-0.0252 \pm 0.0011$ ( $p < 0.0001$ )  35–65, logit gradient = $0.0130 \pm 0.0056$ ( $p < 0.0001$ )	Logit UK < Australia, Middle East, Rest of Europe, Rest of World, South America ( $p = 0.021, 0.045, < 0.0001, < 0.0001, < 0.0001$ ) Logit Australia < Rest of World ( $p = 0.043$ ) Logit North America < Rest of Europe, Rest of World, South America ( $p = 0.01, 0.0001, 0.007$ ) Logit Rest of Europe < Rest of World ( $p = 0.017$ )
4. Reverberated Whoopee cushion	F ( $p = 0.0001$ )	5–35, logit gradient = $-0.0199 \pm 0.0099$ ( $p < 0.0001$ ) 35–65, logit gradient = $0.0160 \pm 0.0054$ ( $p < 0.0001$ )	Logit UK < North America, Rest of Europe ( $p < 0.0001, 0.001, p < 0.0001$ ) Logit Rest of World > South America ( $p = 0.032$ )

The column gender indicates whether Females (F) or Males (M) found the sound more horrible; ‘–’ indicates no statistically significant difference.

Sounds 2 and 4 were the same recording of a whoopee cushion, and was meant to be like the sound of breaking wind. The difference between the recordings is that sound 4 had a considerable amount of reverberation added to the signal. This significantly changed the sound, as evidenced by the fact that one lies near the top and one near the bottom of the rank ordering. Both sounds gave a similar variation with age, a U-shaped response with the youngest and oldest respondents finding the noises more horrible than those in the middle age group. These sounds might be comic as well as horrible and maybe that can explain the response variation with age. Sound 2 was rated similarly

by both genders, the addition of reverberation to make Sound 4 resulted in females finding the sound worse. Sound 2 was rated differently between many locations; there were fewer significant variations with location for Sound 4.

## 5.2. Discussions

At the outset of the experiment, it was expected that the results from the disgusting sounds would follow the clear pattern found by others looking at disgust using visual stimuli. However, the results indicate this is not the case. This might be due to the methodology. The



experiment here did not directly ask about disgust but about horribleness which allowed respondents to bring in other issues when deciding on a rating. Most of these sounds are, however, clearly identifiable and so it is highly likely that it is the association with the source, e.g. someone coughing, that is the cause of the unpleasant reaction. Consequently, this suggests the response to the sound is not just about disgust and disease avoidance, but other factors are involved. The most likely factors relate to social disgust and whether it is acceptable to make disgusting sounds in public. There is a difference between aural and visual stimuli. If one encounters something that looks disgusting, say someone who looks ill, it is often possible to avert one's eyes and thereby remove the unpleasant stimulus. If the ill person is creating noise, however, say by coughing, it is much more difficult to remove the unpleasant stimulus. The sound of something disgusting is often more invasive than the sight of something disgusting. Maybe the more invasive nature of disgusting sounds has meant that stronger social disgust reactions have developed. This does not mean that disease avoidance is not important because the intrinsic dislike of disgusting acts for reasons of survival may have been strongly modified by cultural factors. However, to understand this better requires further experimentation where the reasons behind the disgusting responses are tested in detail.

## 6. Conclusions

A range of horrible sounds have been examined in a web-based psychoacoustic experiment. By using the Internet, it was possible to have tens of thousands of people auditioning and rating sounds, however, the data gathered is inherently 'noisier'. The results from two categories of noises: scraping and disgusting sounds are presented in this paper. Scraping sounds are interesting because it includes the archetypal worst noise – the sound of fingernails scraping down a blackboard. Disgusting sounds are investigated because although much work has previously been done on disgust, the audio aspects have previously been rather neglected.

Scraping sounds can invoke strong reactions, yet the underlying reasons for this reaction is not understood. 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 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 result supporting the hypothesis, and the other not. This indicated that the audio–haptic concept warrants further investigation. Two

new hypotheses for the response to this sound were forwarded drawing on work concerning dissonance; further work is needed to examine these.

A set of eight noises that were expected to invoke a disgust reaction were used in the experiment. These included the worst found, the sound of someone vomiting. However, none of the sounds provided responses consistent with a disgust reaction related just to disease avoidance and survival. This is evidence that other factors influence responses to these horrible sounds, with socially learnt disgust being suggested as the most obvious factor.

Overall, the web experiment has produced a wealth of interesting experimental results on how our response to horrible sounds varies with gender, location and age, some of which are presented in this paper. However, the experiment has probably raised as many questions as it answered. Detailed laboratory studies which ask people why they give a sound a particular rating are needed to enable a more thorough understanding of the voting patterns. This laboratory experimentation is also needed to test the robustness of carrying out psychoacoustic testing across the Internet.

## Acknowledgements

The work was funded by the Engineering and Physical Research Council (UK). Thanks to Prof. Rose Baker for advising on appropriate statistical analysis methods.

## References

- [1] Cox TJ. The effect of visual stimuli on the horribleness of awful sounds. *Appl Acoust* 2008;69(8):691–703.
- [2] Reips U-D. Standards for internet-based experimenting. *Exp Psychol* 2002;49(4):243–56.
- [3] Cox TJ. Bad vibes: an investigation into the worst sounds in the world. In: *Proceedings of the 19th ICA Madrid*, PPA-09-003, 2007.
- [4] Conover WJ. *Practical nonparametric statistics*, third ed. New York: John Wiley & Sons; 1999.
- [5] Agresti A. *An introduction to categorical data analysis*. Wiley series in probability and statistics, 1996.
- [6] Halpern DL, Blake R, Hillenbrand J. Psychoacoustics of a chilling sound. *Percept. Psychophys.* 1986;39(2):77–80.
- [7] Blake R. Margins – primal screech. *Psychol Today* 1986;20(9):68.
- [8] McDermott J, Hauser M. Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. *Cognition* 2004;94(2):B11–21.
- [9] Curtis V, Aunger R, Rabie T. Evidence that disgust evolved to protect from risk of disease. *Proc. R. Soc. Lond. B* 2004(Suppl.271):S131–3.
- [10] Fritz, Thomas. *Personnel communication*.
- [11] Gaver WW. What in the world do we hear? An ecological approach to auditory event perception. *Ecol Psychol* 1993;5(1):1–29.
- [12] Dubois D. Categories as acts of meaning: the case of categories in olfaction and audition. *Cogn Sci Quart* 2000;1:35–68.
- [13] Czigler Istvan, Cox TrevorJ, Gyimesi Kinga. Event-related potential study to aversive auditory stimuli. *Neurosci Lett* 2007;420(3):251–6.
- [14] Kamo M, Iwasa Y. Evolution of preference for consonances as a by-product. *Evolution Ecol Res* 2000;2(3):375–83.
- [15] McDermott J, Hauser MD. Probing the evolutionary origins of music perception. *Neurosciences and music ii: from perception to*

- performance annals of the New York academy of sciences, vol. 1060. 2005;6–16.
- [16] Huron D. A possible evolutionary origin for auditory dissonance. CCRMA hearing seminar 1997.
- [17] Curtis VA, Biran A. Dirt, disgust and disease: is hygiene in our genes? *Perspect Biol Med* 2001;44:17–31.
- [18] Rozin P, Haidt J, McCauley C. Disgust. In: Lewis M, Havilandand JM, editors. *Handbook of emotions*. New York: Guilford Press; 2001. p. 637–53.
- [19] Rozin P, Fallon AE. A perspective on disgust. *Psychol Rev* 1987;94(1):23–41.