NTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

This is an edited version of Prof. Cox's original paper.

Student Activity Sheet 1 Peer Review

BAD VIBES: AN INVESTIGATION INTO THE WORST SOUNDS IN THE WORLD

PACS: 43.66.Lj

Cox, Trevor J.
[1]Acoustics Research Centre, University of Salford, Salford M5 4WT, United Kingdom; t.j.cox@salford.ac.uk

ABSTRACT

An Internet experiment has been carried out where people listen to and rate a variety of horrible sounds such as the sound of: fingernails being scraped down a blackboard; someone throwing up, and a dentist drill. Although the experimental conditions are very uncontrolled, using the Internet allows access to a vast number of ratings (1.5 million to date) and to a range of subjects that is virtually impossible to do in the laboratory. This paper will explore the methodology, examining the pitfalls and issues with running such an experiment, for instance the problems arising from the lack of calibration and uncontrolled context. The experiment works well in examining how responses to sounds vary with contextual variables such as age, country and gender. The website was also used to examine if varying the colour of the website affected ratings. Some of the interesting results from the experiment are explored.

INTRODUCTION

A key part of acoustic science is to understand how humans respond to sound, and because of this, there are many different experimental techniques that can be used. Most perceptual experiments rely on small sample sizes, because the process of testing a large number of subjects is too time consuming. It is now possible, however, to access a large worldwide audience via the Internet.

Testing human responses to stimuli via the Internet has become increasingly popular in many areas of psychology. Spanning diverse subjects such as beauty [1], experimenters have been exploring the use of the Internet as a vehicle for experimental psychology. While a methodology which uses the Internet has difficulties, and these will be discussed in more detail below, the opportunity to test a large number of people across the world is highly appealing. To answer some questions in psychology, especially ones relating to culture, it is necessary to test responses from subjects of all ages and people who come from different cultures; it is usually impracticable to test such a wide range of people in a laboratory experiment.

In this paper, interest lies in experiments using the general public mediated through the Internet.

Nowadays, nearly all computers are sold with the ability to reproduce sound and to connect to the Internet. The number of broadband connections is rapidly increasing. Consequently, it is now possible to carry out tests across the Internet where people actually listen to and judge sounds. The purpose of this paper is to describe experience of running such an experiment, to evaluate the methodology and highlight the advantages and disadvantages in carrying out experiments across the Internet.

METHODOLOGY

Subject recruitment

Given the lack of control over the methodology, for the volume measurement at which people listen to the sounds, Internet experimentation with sounds is usually only going to produce statistically significant results if a large dataset (number of results) is gathered, so the errors due to the uncontrolled effects can be ‘averaged’ out. So we want to access a very large number of people. In this context, a large number is tens of thousands of subjects. Efforts to gain this large number of subjects impinges not only on methods for recruiting subjects, but also on what psychological effects of sound on people can be tested through the methodology.

Conventional media, newspaper, television and radio were found to be good and reliable ways of publicising the web experiment. While word-of-mouth can lead to websites becoming very popular, it is unlikely that an experiment will gain a large enough number of subjects by word-of-mouth alone. Furthermore, it is immensely difficult to forecast what websites will become popular. So it is risky to set up an experiment and just hope people will find it. Therefore, the web experiment must be about a subject that will appeal to the media. It helps if

the web experiment has a serious scientific purpose with a good news angle. For the national media, it is likely that the story will be dealt with by a science journalist, and so they will need to be reassured that there is a proper scientific purpose behind the work. A good news story, of interest to readers, is needed. If the web experiment is quirky or unusual then not only is it possible to gain media coverage, but it is more likely that word-of-mouth will get more people to the experiment. For these reasons, the Internet experiment recently carried out was to search for the Worst Sound in the World which at the time of writing has attracted 1.5 million ratings, from about 130,000 voters.

Figure 1.- An eye-catching image used to sell story to the press

It is vital that the website is hosted on a reliable hosting service with sufficient bandwidth to cope with sudden increase in users that this publicity can generate. Mass Internet experiments are one-off experiments. It would be difficult to repeat the experiment again because the media will not be interested. Consequently, it is important that the experimental design is thorough and carefully considered and hosting is robust, because repetition is not possible.

Experimental design

The need to sell the experiment to the media influenced the experimental design. It was sold to the media as being the search for the Worst Sound in the World, which naturally meant the website had to have a large number and range of sounds. 34 sounds were used. However, this meant the experiment was rather unfocussed and made the scientific analysis more difficult. The media were very interested in the idea of finding the worst sound in the World, but in many ways the rank order of sounds produced was rather meaningless. The rank order is dependent on factors such as the quality of recordings and the typical context of listeners. However, on the back of this media-friendly concept it was possible to make more reliable assessment of relative changes, for instance what happens when the website changed colour, and this is where the scientific experimentation became useful and more valid.

The experiment described here was aimed at the general public. Consequently, the method had to be simple, east to use 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 some contextual data to be able to interpret the voting patterns.
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 acoustic experiments on the web is the lack of control over the loudness of the sound reproduced. In this type of experiment, 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 measurement of volume at which people listened to the sounds 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 by getting sufficient numbers of people, 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 context variables, say age, then this could also introduce a bias into the results. However, this bias can be minimized by using appropriate analysis. Overall, the lack of measurement of volume at which people listen to the sounds is a significant drawback to carrying out acoustic experiments on the web, and tests need to be carried out to compare results from web and laboratory experiments to test whether the ‘averaging’ out of error is correct.

Next, participants came to the voting screen which can be seen in Figure 2. Users pressed play, listened to the sound and then voted using the smileys at the bottom of the screen. A key to making web experiments popular is to make the interface simple, interesting and fun to use, and this is why smileys were chosen. Consequently, this is a direct scaling method. It would have been possible to use a slider scale and therefore obtain scale data, however the use of the visually-appealing smileys was preferred as being less difficult to use.

Figure 2.- The voting screen used for the website.

The box in the middle of the screen might have displayed an image or just have been blank with a simple text message. After voting, a results page was given which showed the user's own response and compared this to the average response for that sound. It was felt important to give users some feedback from their votes, but there is a risk that this feedback might bias future ratings. 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 analyse how many votes had been cast by each user.

Users could continue to listen to as many sounds as they wanted; there were 34 in the database. With the benefit of hindsight, this was not the best approach. The most common email from users was to ask how many sound files there were to listen to. Consequently, it would have been better to get users to vote on a fixed number of sounds. This would also have helped in analysing the data. On average people listened to 11.5 sounds. People were given the chance to vote at anytime while listening to the sound file. It is suspected that this meant sounds which were horrible from the beginning of the sound file 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.

Discussion

Using the web to run experiments offers a number of advantages, but also a series of methodological challenges [2]. As noted before there is no measuemrnt of the volume at which the sounds were listened to, and quality of

the speakers varies between people. People are self-selecting and the place in which they listen to the sounds is uncontrolled.

SOME RESULTS

During the experiment, what was on the screen was varied. During the first few weeks of the experiment, an image might have been displayed; later on the colour of the website was varied. This allowed the interaction between sound and vision to be explored. This produced a host of interesting results discussed in detail elsewhere [3]. Probably the most interesting result from this part of the study, was the finding that sounds were rated to be more horrible for websites with more relaxing colour schemes. A lesson learnt from this part of the work was if changes are going to be made to the website such as altering colour schemes, these changes need to be designed, implemented and tested in advance of launching the website. Some of the work on images was implemented on-the-fly, and while no problems were encountered, this is poor experimental practice. Ultimately, the work on images showed that significant differences in rating could be found with 100,000 votes cast across the 34 sounds. However, developing a meaningful and reliable interpretation of the results in many cases was difficult.

It is important to find out whether it is best to tell listeners what the sound is, or to get them to listen to the sound unidentified. For example, it seems that many people think the sound of fingernails dragging down the blackboard is the worst sound in the World, yet the actual recording of the sound only came midway down the ranking list. When an image of fingernails on a blackboard were shown while the sound was being listened to, then the sound was rated much more horrible. This effect was seen in many cultures [5]. It is known that in describing sounds, people attempt to assign interpretations as to the event or source creating the sound [6-7]. Consequently, what people are told is important; this is also true of laboratory experiments.

For a fixed visual stimulus, the variation in responses with age, gender and location can be examined. Variations with age and gender are potentially easier to interpret than variations with location. To illustrate this, consider the sound of the dentist drill. This was significantly more horrible for people in South America and significantly less horrible for those in Africa. This presumably says something about the quality of dentistry in South America and maybe the likelihood of having dentistry in Africa. Without a more detailed study, however, and an understanding of dentistry practises across the world, it is rather hard to do draw any useful conclusion from such a result. To take another example, compared to listeners elsewhere in the World, Australians rated the soap opera argument to be significantly more horrible. Maybe this has something to do with the Australians' laid-back attitude to life, but this is pure speculation without further research.

The results for scraping sounds have been analysed [4]. It is not understood why some humans find certain sounds, such as the sound of fingernails being scraped down a blackboard, so terrible. Some people have strong reactions to this sound, for instance covering their ears or even leaving the room.

CONCLUSIONS

The Internet offers an interesting additional technique for experiments which look at the effect of sounds on people. The experiments give access to large datasets with people drawn from different backgrounds, but only for experiments that are media-friendly. The experimental conditions become rather uncontrolled. There is need for more research to examine the validity of the method, as has been done for Internet experiments which have not used sounds as stimuli. A set of laboratory experiments is needed to examine whether the results from the Internet experiment are reliable.

References

[1] J. H. Krantz, J. Ballard and J. Scher, “Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness,” Behavior Research Methods Instruments & Computers 29(2), 264-269, (1997).
[2] U.-D. Reips, "Standards for internet-based experimenting," Experimental Psychology, 49(4), 243-56, (2002).
[3] T. J. Cox, “The effect of visual stimuli on the horribleness of awful sounds,” Applied Acoustics, accepted for publication (2007).
[4] T. J. Cox, “Scraping Sounds and Disgusting Noises,” submitted (2007)
[5] W. W. Gaver, “What in the World Do We Hear? An Ecological Approach to Auditory Event Perception,” Ecological Psychology, 5(1), 1-29, (1993).
[6] D. Dubois, “Categories as Acts of Meaning: The Case of Categories in Olfaction and Audition,” Cognitive Science Quarterly 1, 35-68, (2000).