Effects of very high-frequency sound and ultrasound on humans. Part II: A double-blind randomized provocation study of inaudible 20-kHz ultrasound

Mark D. Fletcher, Sian Lloyd Jones, Paul R. White, Craig N. Dolder, Timothy G. Leighton, and Benjamin Lineton

Citation: The Journal of the Acoustical Society of America **144**, 2521 (2018); doi: 10.1121/1.5063818
View online: https://doi.org/10.1121/1.5063818
View Table of Contents: http://asa.scitation.org/toc/jas/144/4
Published by the Acoustical Society of America

Effects of very high-frequency sound and ultrasound on humans. Part II: A double-blind randomized provocation study of inaudible 20-kHz ultrasound

Mark D. Fletcher,[1,a)] Sian Lloyd Jones,[2] Paul R. White,[1] Craig N. Dolder,[1] Timothy G. Leighton,[1] and Benjamin Lineton[1]

[1]*Faculty of Engineering and Physical Sciences, University of Southampton, University Road, Southampton SO17 1BJ, United Kingdom*

[2]*Department of Audiology and Hearing Therapy, Royal South Hants Hospital, Brinton's Terrace, Southampton SO14 0YG, United Kingdom*

(Received 7 August 2017; revised 23 January 2018; accepted 23 January 2018; published online 31 October 2018)

Some people have reported symptoms such as nausea, dizziness, and headaches that they attribute to ultrasound (US) emitted by devices in public places. The primary aim of the present study was to investigate whether inaudible US can provoke adverse symptoms compared to a sham presentation, under double-blind conditions. A second aim was to investigate whether the expectation of US being present could provoke adverse symptoms (a nocebo response). The US stimulus was a 20 kHz tone presented continuously for 20 min set to at least 15 dB below the participants' detection threshold, giving a typical sound pressure level (SPL) of 84 dB. No evidence that US provoked symptoms was found, but there was evidence of small nocebo effects. A case study on an individual with high self-reported sensitivity to US gave similar results. The present study did not reproduce the severe symptoms reported previously by some members of the public; this may be due to the SPL or duration of the stimulus, or strength of the nocebo stimulus. These findings cannot be used to predict outcomes from exposures to sounds that are audible to the individual in question, or to sounds with higher SPLs, longer durations, or different frequency content. © *2018 Acoustical Society of America*.
https://doi.org/10.1121/1.5063818

[JFL] Pages: 2521–2531

I. INTRODUCTION

Since the late 1940s, there have been complaints of symptoms experienced by workers exposed to ultrasound (US) from tools and machinery.[1] These symptoms include nausea, headaches, dizziness, tingling in the limbs, and fullness in the ears (reviewed by Lawton, 2001; Leighton, 2016). One of the first documented attributions of such symptoms to US exposure came from air force personnel working with early jet engines (Pharris, 1948). These initial media reports were met with caution by the scientific community because of their anecdotal nature and because ultrasonic exposure was accompanied by high level audio-frequency exposure (Leighton, 2016). Since that time, several studies have recorded cases of similar symptoms experienced by workers in the industry using machinery, such as ultrasonic cleaners (Acton and Carson, 1967; Crabtree and Forshaw, 1977) and welders (Macca *et al.*, 2014), although again with accompanying audio-frequency exposure. In recent years, the variety of ways in which workers and the public can be exposed to US has increased, including tonal US exposure without intense audio-frequency content in public settings (Leighton, 2007). Many ultrasonic devices, such as pest deterrents and public address voice alarm (PAVA) monitoring systems, have been deployed in public places, thereby exposing members of the public, often without their knowledge, to US (Leighton, 2016, 2017). In Part I of this paper, we reported evidence of symptoms produced by audible very high-frequency sound (VHFS; Fletcher *et al.*, 2018a). There have been some anecdotal reports conveyed to our research group and to the national press by members of the public that adverse symptoms can arise from exposure to inaudible US produced by such devices (e.g., Ebelthite, 2016; Fletcher, 2016). The presence of US in air in some of the public places identified has subsequently been confirmed (Leighton, 2016). Because there is yet to be an adequate study to confirm whether or not a causal association exists between exposure to US and the physiology underlying the reported symptoms, indirect causes must be considered. These might include anxiety as a consequence of being alerted to the presence of inaudible US either by seeing a source such as a pest deterrent, or by measuring the sound field using a smartphone, which is able to detect sounds up to ~22 kHz (Leighton, 2016, 2017). To date, there have been no double-blind controlled trials to test for the effects of inaudible US.

The present study is a double-blind trial exploring the effects of inaudible US at sound pressure levels (SPLs) that might be encountered by the general public. Two groups of participants were studied; those who reported adverse symptoms that they attributed to very high-frequency sound (VHFS) or US (termed here the "symptomatic" group) and those who did not (the "asymptomatic" group). Symptoms that participants in the symptomatic group attributed to exposure from VHFS/US included nausea, pain in the ears, headache, pressure in the ears or head, dizziness, anxiety, annoyance, tiredness, and inability to concentrate.

[a)]Electronic mail: M.D.Fletcher@soton.ac.uk

The main objective of the study was to determine whether exposure of 20 min to inaudible US leads to adverse symptoms in either asymptomatic or symptomatic participants, when compared to a sham exposure control condition (i.e., no exposure). A secondary objective was to determine whether there was a nocebo effect arising entirely from the participant's belief that US was present.

It is conceivable that US may not be detectable during hearing threshold testing, when the stimulus duration is short, but can nevertheless be inferred by the participant in the main experiment, when the stimulus is much longer, because of the emergence of associated symptoms. For this reason, the ability of participants to correctly judge whether inaudible US is present under double-blind condition is also assessed in this study.

The SPL and duration of the stimulus was strictly regulated by the ethical permissions received, and was significantly less than members of the public might receive from some devices (e.g., close to some pest deterrents). The maximum SPLs allowed for the 20 min exposure to a 20 kHz tone that was used in this study was 88 dB SPL (all SPLs stated re 20 μPa). To ensure that the tone was inaudible, the level was set to at least 15 dB below the hearing threshold for each individual. Following completion of the main experiment, a case study was conducted on a participant from the symptomatic group who reported the most extreme and persistent symptoms resulting from exposure to US in public places. In the case study, the experimental tone was again presented at 20 kHz, but the level was increased to 94 dB SPL (still at least 10 dB below the participant's threshold). All conditions were repeated 12 times, for a period of 1.5 min each.

II. METHODS

A. Procedure

1. Experimental session structure

The structure of an experimental session is shown schematically in Fig. 1. Participants were first assigned the status of either symptomatic or asymptomatic regarding their self-reported sensitivity to VHFS/US, whether audible or inaudible, and screened according to the criteria described in Sec. II B. The same experimental procedure was conducted for both groups of participants.

After screening was complete, the Galvanic skin response (GSR) electrodes were attached to the participant's fingers. Hearing threshold levels at 20 kHz were then measured using a three-interval three-alternative forced choice (3I3AFC) procedure (Sec. II A 2). Hearing thresholds at very high or ultrasonic frequencies are known to vary with removal and replacement of the headphones (e.g., Stelmachowicz *et al.*, 1989) and therefore once the headphones were positioned over the participant's ears, they were not removed for the duration of session.

After the hearing threshold for the 20 kHz tone that would be used in the exposure condition had been measured, baseline ratings and baseline GSR were measured. The GSR baseline measurement lasted 2 min, with the average of the last 1 min used as the baseline. Participants were next familiarized with the sustained attention to response task (SART; Sec. II A 3) in a practice session before the test condition began, to ensure they achieved a performance level of less than 10% omission errors (not pressing when they should) and 50% commission errors (pressing when they should not). If the participant did not achieve the required performance level, the practice trial was repeated (with task instructions reiterated on the screen) to ensure that the participant understood the task. A break of 40 s was then given to allow the GSR to return to baseline. The test conditions, described in Sec. II A 6, then began and ran continuously for 20 min. During this period, the participants carried out the SART for periods of 4 min, followed by a 1 min period where they completed the subjective ratings. When completing the subjective ratings, participants were shown a countdown of the time remaining before the experiment automatically moved on, and were instructed to complete all ratings before this time elapsed. This was repeated four times (covering the total exposure time of 20 min). In each of the double-blind test sessions, after the exposure or sham exposure was complete, participants were asked whether they thought US was present (yes/no) and to rate their level of certainty (Sec. II A 5).

At the end of the session, the participant's hearing threshold at 20 kHz was again measured to ensure that the sensation level had not changed significantly, for example as a result of changes in headphone position. The unsigned average difference between the pre-and post-hearing threshold levels was 1 [$\pm$0.6 (standard error of the mean)] dB for the symptomatic group and 1.3 ($\pm$0.3) dB for the asymptomatic group. In no case did the estimates of pre- and post-experiment threshold for the exposure condition imply sensation levels used for the exposure condition were greater than $-$12.5 dB. It should be noted that this threshold stability is partly due to the fact that, for many participants, the 20 kHz tone was inaudible, even at the maximum tested SPL of 105 SPL.

A single case study was also conducted on a participant who reported particularly strong symptoms resulting from US exposure in public places. The methodology for this case study is presented in Sec. II A 9.

2. Estimation of pure-tone hearing thresholds

Throughout this paper, hearing thresholds are expressed in SPL rather than in hearing level (HL), since there is no reference equivalent threshold SPL at 20 kHz. To ensure that participants could not detect the presence of the US via a hearing percept, it was necessary first to measure their hearing threshold at the exposure frequency of 20 kHz.

Hearing thresholds were measured using an automated 3I3AFC paradigm controlled by the laptop in the observation room. In these hearing threshold measurements and in all subsequent tests, the stimulus was presented diotically. Each trial comprised three listening intervals. One interval, chosen randomly with equal *a priori* probability, contained the signal, and the other two contained silence. The participant selected the interval that they thought contained the signal using a mouse. Visual feedback was given after each trial indicating whether the response was correct or incorrect. The listening intervals were 550 ms in duration, separated by 300 ms of silence. The stimulus steady-state duration was

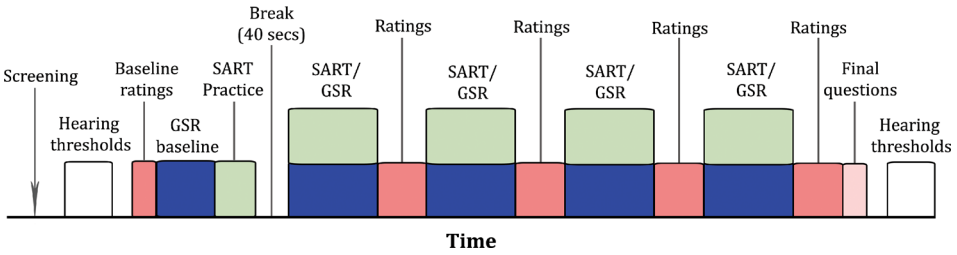

FIG. 1. (Color online) Schematic (not to scale), showing the timeline of a single session.

500 ms, with 25 ms quarter-sine and quarter-cosine ramps at the beginning and end of the stimulus (making the total duration between 0 volt points 550 ms).

The stimulus SPL was varied using a two-down, one-up procedure: the stimulus SPL was increased for a single incorrect response, and reduced following two consecutive correct responses. The stimulus SPL was initially set at 90 dB, and was changed in 10 dB steps up to the first reversal, in 5 dB steps up to the second reversal, and in 2.5 dB steps for the remaining eight reversals. The threshold was defined as the mean of the final eight reversals. The staircase terminated early if there were three incorrect responses at the maximum stimulus SPL of 105 dB and the hearing threshold was assumed to be above this level. This procedure estimates the SPL at which the participant can correctly detect the target 70.7% of the time (Levitt, 1971).

3. Performance on a sustained attention task: Sustained attention to response task (SART)

The SART was used to assess lapses in concentration that could arise from exposure to inaudible US (see Robertson *et al.*, 1997; Manly *et al.*, 1999; Manly *et al.*, 2002; Manly *et al.*, 2003; Smilek *et al.*, 2010; Foxe *et al.*, 2012). In the SART, digits from 1 to 9 were displayed on a computer screen in sequence, with the participant required to click a mouse button when any number appeared apart from the number 3 (the target). The target (the number 3) appeared in 10% of trials (selected at random). Digits were displayed for 150 ms after which no stimulus was displayed for a duration of between 1000 and 1500 ms (randomly varied). Participant responses were recorded from the time at which the digit appeared to the end of the trial (i.e., any time before the next digit was displayed). SARTs lasting 4 min were completed at four sequential time points for each condition (Sec. II A 6), with 1 min between each when rating scales were completed. A five second countdown was given before the start of each SART so that the start of the trial was not unexpected. Instructions were given before a practice test lasting 45 trials, which was at the start of each of the four separate sessions. Participants were instructed to give equal priority to speed and accuracy in their responses.

4. Galvanic skin response

As for the SART task, GSRs were measured at four sequential time points for periods of 4 min. The GSR measurements within each session were subtracted from a baseline condition measured at the start of each session.

5. Subjective ratings of symptoms

At four times during the experiment, the participant was asked to give a subjective rating of the severity of the following 10 items: overall discomfort, nausea, pain, pressure, or fullness in one or both ears, headache/pain or pressure somewhere other than the ears, dizziness or lightheadedness, tinnitus, anxiety, fatigue, and other symptoms. The format of questions posed to participants is given in Appendix B. Overall discomfort was always rated first and "other symptoms" was always rated last, but all other symptoms were rated in a random order each time that ratings were given. Participants were asked to give a rating for each item on an 11-point scale from 0 to 10, with "0" and "10" given the descriptors of "not at all" and "severe," respectively. These items were chosen as ones that have been previously associated with exposure to VHFS/US (Skillern, 1965; Acton and Carson, 1967; Acton, 1974; Crabtree and Forshaw, 1977; Herman and Powell, 1981; Acton, 1983; Maccà *et al.*, 2014; Ueda *et al.*, 2014), or had been reported by respondents to the recruitment material. The "overall discomfort" item was included as the primary outcome measure to allow comparisons across participants who may experience diverse symptoms, or may select different descriptions of potentially similar symptoms.

At the start of each session, participants were screened to ensure they had not been experiencing symptoms (see Appendix B). For this screening, the prefix "Over the past hour…" was used in order to ensure that symptoms had not recently been present and temporarily subsided. In each session, baseline ratings of symptoms were also collected after the threshold measurements had been completed, just before the main experiment began. These ratings were subtracted from all other symptom ratings given within the session (for these ratings, the prefix "Over the last 4 min…" was used).

At the end of the test conditions in which the participant was told that US may or may not be present (see Sec. II C 6), participants were also asked to state whether they thought US was being presented ("yes" or "no") and to rate their confidence in their response to this question. The confidence ratings ranged from 0 ("totally uncertain") to 10 ("totally certain").

6. Test conditions

Each asymptomatic participant underwent four test sessions, separated by a minimum of 2 h and taking place over a minimum of two days. In these sessions, two independent variables were manipulated. The first independent variable was the US presentation status which was either present or

TABLE I. Experimental conditions.

Condition identifier	Shorthand descriptor	US State	Information to participant
A	US on; No cue	On	US possibly present
B	US off; No cue	Off	US possibly present
X	US off; False cue	Off	US on
Y	US off; True cue	Off	US off

TABLE II. Order of testing.

	Session			
Participant arm	*1*	*2*	*3*	*4*
Arm 1	*A*	*X*	*B*	*Y*
Arm 2	*A*	*Y*	*B*	*X*
Arm 3	*B*	*X*	*A*	*Y*
Arm 4	*B*	*Y*	*A*	*X*

absent (i.e., no exposure). During both of these conditions the participant was told (truthfully) that US may or may not be present and that the tester did not know whether or not US was present. The comparison of these two conditions allows the effect of exposure to US to be determined. During the US present condition, US was presented at a frequency of 20 kHz, and to an SPL set to either 88 dB SPL or to 15 dB below their hearing threshold level (whichever was lower). The average SPL that the signal was presented at for the asymptomatic group was 83.3 dB [±1.4 dB (standard error of the mean), ranging from 57.2–88 dB], and across symptomatic participants was 83.7 dB (±2.7 dB, ranging from 65.6–88 dB). The total duration of the exposure was 20 min, with 25 ms quarter-sine and quarter-cosine ramps at the beginning and end of the stimulus, respectively.

The second independent variable was the participant expectancy cue. This manipulated the participant's expectation of the presence or absence of the US using a display, which indicated either "Ultrasound ON" or "Ultrasound OFF." In both cases, there was no actual US being presented, and thus the first of these conditions was a "false cue" condition, requiring deception of the participant. The second condition was a "true cue" condition. The comparison of these two cue-conditions allows any nocebo effect to be assessed.

At the start of the experiment, the experimenter explained to participants that there would be both "unknown conditions," where neither the experimenter nor the participants would know whether US was present, and "known conditions," where whether US was present or absent would be displayed on the screen. No other details about the experiment design were given. The four conditions are denoted *A*, *B*, *X*, and *Y*, and are summarized in Table I.

Participants were assigned to one of the four arms shown in Table II, to determine the order of testing. This design is effectively two randomized interleaved cross-over trials for the two levels of the exposure condition and two levels of the cue condition. Condition *A* or *B* was always carried out first, followed by either *X* or *Y*. The experiment was always over a minimum of two days, with a maximum of two conditions completed in any one day (separated by at least 2 h). Participants always undertook conditions *A* and *B* and conditions *X* and *Y* at the same time of day (morning, afternoon, or evening). This means that any diurnal effects were minimized for comparisons of conditions *A* with *B*, and *X* with *Y*.

Each of the four sessions lasted for around 30 min. Each exposure or sham exposure period (of 20 min) was separated into four parts. In each part, the SART was performed and the GSR recorded for 4 min, followed by 1 min for participants to provide subjective symptom ratings before the experiment automatically continued to the next part. At the end of the sessions for conditions *A* and *B*, participants were asked whether or not they thought US was present, and what their confidence in this answer was. During all conditions a stop button was displayed on the screen. This allowed the participant immediately to stop the experiment at any time. No participants stopped the experiment.

7. Double-blinding method and concealment of false-cue from participants

To maintain blinding of the researchers and participants to whether condition *A* or *B* was used for a given test session, participants were assigned to one of the arms in Table II by a computer algorithm that also ensured counterbalancing to give equal numbers of participants for each arm. Blinding was maintained throughout the experiment and data analysis, with unblinding only occurring immediately prior to submission to the journal for review.

Researchers were not blind to conditions *X* and *Y*, as they could see the cue given to the participant. All participants were made aware of the use of deception in condition *X* during a debriefing session that occurred after the final participant had been tested.

8. Sample size calculation

For both groups, the primary outcome measure was the difference in overall discomfort rating between the genuine and sham exposure conditions, where no cue was presented (conditions *A* and *B* in Table I). An increase in rating of three points would be considered important, as it represents an increase from no noticeable effect to a clearly noticeable adverse effect. A previous pilot study indicated that the measurement test-retest error (standard deviation of measurements in identical exposure conditions) to be typically <one point in asymptomatic participants who typically scored near the bottom of the rating scale throughout the test. However, this is likely to be an underestimate for cases where symptoms arise. For the purposes of this study, the target sample size was set using the much more conservative estimate of the standard deviation of six points.

The pre-test sample-size calculation was based on the directional alternative hypothesis. This was, specifically, that the effect of US in condition *A* would be a three point increase on the response scale for the primary outcome measure compared to the sham exposure to US in condition *B*. The standard deviation of the difference measure was assumed to be six points. For a type 1 error rate of 5%, and a

power of 80%, this gave a required sample size of 27 for a paired t-test. Although the distribution of the outcome measure was known to be positively skewed, violating the assumption of normality required for the t-test, the test was deemed appropriate given the conservative assumptions regarding standard deviation.

This sample size was exceeded in the asymptomatic group (32 participants), though not in the symptomatic group (eight participants).

9. Additional case study

In an additional case study, repeated shortened test conditions were run using a higher SPL than in the main experiment. The participant who took part in this case study had reported particularly strong effects of exposure to US in public places, based on responses to the pre-existing symptom questionnaire (Appendix A). They attended a single additional session, in which they completed all four arms shown in Table II three times, meaning they completed 48 test conditions in total (12 repeats of each condition). Each test condition was reduced in time from 20 to 2.5 min. The order in which the arms were completed was chosen at random. The participant carried out the SART for 1.5 min from the start of each condition with GSR being measured concurrently, during which time the actual, sham exposure, or cued conditions were active. This was followed by a period of ~1 min when the subjective ratings were completed, before the next condition began. A 10 min break was given at the halfway point, in which the participant was allowed to remove the headphones. Hearing thresholds for the 20 kHz tone were measured before and after the experiment, and in both cases were >105 dB SPL (at least 10 dB above the SPL of the 20 kHz tone). After each of the blinded conditions, the participant was asked whether they thought US was present ("yes" or "no"). In total, the testing session lasted ~2.5 h.

B. Participants

Participants were split into two groups based on their responses to a questionnaire, which assessed whether they had experienced symptoms that they attributed to exposure to VHFS/US. Eight participants (four females and four males, with an average age of 28 yr, ranging between 20 and 40 yr) who had experienced symptoms, were recruited to a "symptomatic" group. Owing to difficulties in recruiting participants who reported symptoms that they attributed to exposure to US (chiefly because the experiment ran over multiple days, making it impractical to recruit participants based far from the University of Southampton), participants were categorized as symptomatic if they attributed symptoms, including, ear pain, ear pressure, headache, tinnitus, dizziness, or nausea to exposure to audible VHFS/US (six of the eight participants), or to inaudible US (two of the eight participants). Symptomatic participants had either responded to a call for participants who experienced some adverse symptoms that they attributed to exposure to VHFS/US, or had contacted the researchers as a result of their symptoms. The call was put out in the form of posters placed around the University of Southampton, and through social media (Fletcher, 2016).

The group labelled as "asymptomatic" comprised participants who reported no adverse symptoms that they attributed to exposure to VHFS/US, whether audible or inaudible. Thirty-two participants (15 female, average age of 24 yr, ranging between 18 and 33 yr) were recruited to this group.

The exclusion criteria for both groups were troublesome tinnitus, hyperacusis in response to "everyday" sounds, or hearing threshold levels that exceeded 20 dB HL at the standard audiometric frequencies of 0.25, 0.5, 1, 2, 4, and 8 kHz in either ear. Participants were not allowed to take part in a session if they indicated on the screening questionnaire that they had drunk more than one cup of tea or coffee, drunk an energy drink, ingested a pro-plus tablet on the day of testing, had taken recreational drugs in the week leading up to any session, had drunk more than six units of alcohol in the 24 h before the session, or had undertaken strenuous physical or mental activity on the day of testing (see Appendix A). They were also excluded if their subjective rating of symptoms prior to testing in any session exceeded a pre-set threshold of two on an 11-point scale from 0 to 10 (Sec. II A 5).

The experiment was approved by the Ethics Committee of the Faculty and Engineering and the Environment at the University of Southampton (submission number 26450). The study was pre-registered with the Australian New Zealand Clinical Trials Registry (Trial id CTRN12616001410448), but with some aspects of the experimental protocol involving deception of the participants not visible to the public until after the end of the study.

C. Equipment

Participants were seated in a sound-attenuated booth with a background noise level conforming to British Society of Audiology (2017) recommendations. Acoustic stimuli were generated by a laptop located in a separate observation room, and played out via an RME Babyface Pro soundcard (sample rate of 96 kHz and bit depth of 24 bits), a Phonitor 2 headphone amplifier (Sound Performance Lab), and Sennheiser HDA 200 circumaural headphones. The stimuli were calibrated using a Bruel and Kjaer artificial ear (type 4152) with a flat-plate adaptor (DB0843). The two earphones were separated by approximately 145 mm, as specified in ISO 389-5:2006 and the headband tension complied with the requirement of ISO 389-5:2006. The level of any subharmonics measured using the artificial ear was found to be below the noise floor of the artificial ear. Regular checks were made throughout the data collection using this equipment to ensure that the stimulus was being presented at the designated level.

The participants sat facing a computer screen (at a distance of 1 m from the eyes), and operating a computer mouse, both of which were connected to the laptop located in the observation room. The screen and mouse were used for measurements of hearing threshold, for collecting rating scale data, and for administering a visual task used to assess sustained attention.

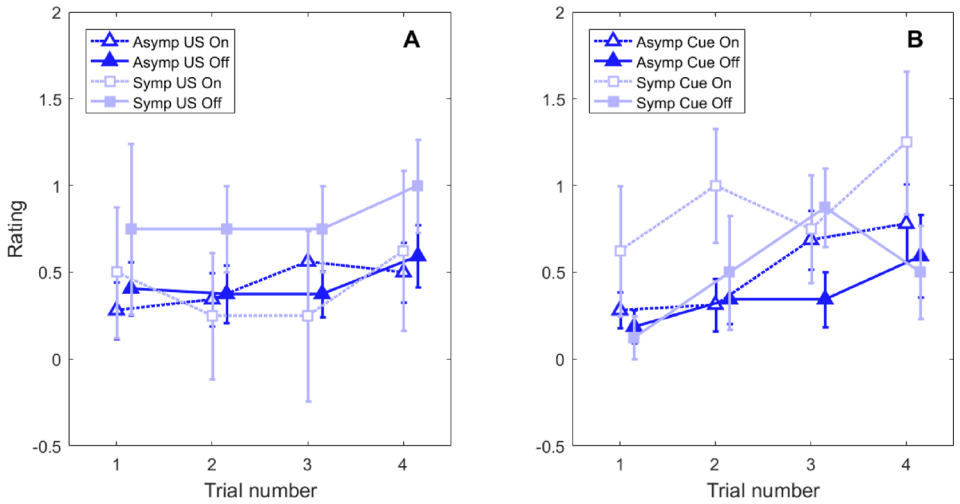

FIG. 2. (Color online) Mean baseline-adjusted discomfort ratings for conditions A and B (panel A) and conditions X and Y (panel B). Means are calculated for each group of participants (32 asymptomatic; eight symptomatic) and are shown for each trial. Markers are horizontally offset for clarity. Error bars show the standard error of the mean across participants.

GSRs were recorded from the proximal phalanges of the index and middle fingers of either the left or right hand (the opposite hand to that which the participant would normally use to operate a mouse), using an Edu-lab Galvanic Skin Response Logger system, which was connected to a second laptop located in the observation room. GSR measures skin conductivity, which varies with changes in the state of sweat glands, and is known to be correlated with psychological state of arousal (Kucera *et al.*, 2004). GSR recordings gave a time-history of skin conductance, with sample rate of 20 Hz.

III. RESULTS AND DISCUSSION

All analysis was completed before any unblinding of conditions *A* and *B* to the researchers.

A. Primary outcome measure

The primary outcome measures were the difference in the overall discomfort rating between conditions *A* and *B* (US on and US off) and conditions *X* and *Y* (the false "on" cue and true "off" cue). Panels A and B of Fig. 2 show the variation in the mean discomfort over the four trials for the US and cue conditions. Baseline ratings taken just before each condition was measured was subtracted from these ratings for each individual.

For both the asymptomatic and symptomatic group, the tendency towards low or zero discomfort ratings led to a strongly positively skewed distribution that approximately followed a gamma distribution. One approach to analyzing the results was to use generalized estimating equations (e.g., Hardin and Hilbe, 2003), which allow a repeated-measures analysis of a gamma-distributed random variable. Using generalized estimating equations to model the overall discomfort rating with stimulus condition as a factor and time as covariate showed no significant effect of either whether US was present [symptomatic group: *Wald* $\chi^2(1) = 0.22$, $p = 0.639$; asymptomatic group: *Wald* $\chi^2(1) = 0.11$, $p = 0.744$] or of cueing (symptomatic group: *Wald* $\chi^2(1) = 0.22$, $p = 0.640$; asymptomatic group: *Wald* $\chi^2(1) = 0.09$, $p = 0.761$) participants for either group.

The statistical results were verified using a non-parametric Friedman's analysis, which also showed no statistically significant increase in overall discomfort [symptomatic group: $Q(1) = 0.5$, $p = 0.43$; asymptomatic group: $Q(1) = 0.2$, $p = 0.32$]. The overall discomfort in the cue conditions also did not reach significance in either group. However, it was close to significance for the asymptomatic group [symptomatic group: $Q(1) = 3.6$, $p = 0.11$, mean effect: 0.4 rating points; asymptomatic group: $Q(1) = 3.2$, $p = 0.074$, mean effect: 0.15 rating points].

B. Secondary outcome measures

The statistical analysis of secondary outcome measures was made without any additional corrections for multiple hypothesis tests. Though less statistically robust than the primary measure, they may nevertheless provide further insight into whether other symptoms were experienced and may indicate whether any strong symptoms might be produced if higher stimulus SPLs, longer stimulus durations, or a stronger nocebo-generating stimuli were to be used in a future study.

1. Subjective ratings

The differences between the genuine US and sham conditions (*A* and *B*) and the two cue conditions (*X* and *Y*) were analyzed for each of the individual items rated by participants, using a non-parametric Friedman's analysis. As for the primary outcome measure, all ratings were baseline corrected. No significant effects of US exposure were found for any symptom (see Fig. 3). Participants were unable to determine whether US was absent or present; across the asymptomatic group, correct responses in both the US exposure and sham conditions were given 21.9% of the time and across the symptomatic group 12.5% of the time. These responses were combined with participants' rating of their confidence in their response to give a score that represented their ability to confidently identify whether US was present. Correct responses for each session were coded as 1 and incorrect responses as −1 and these were multiplied by the confidence rating given [between 0 (totally uncertain) and 10 (totally certain)]. A one-sample Wilcoxon signed-rank test found that these scores were not significantly different from zero (the score expected if participants were guessing) for

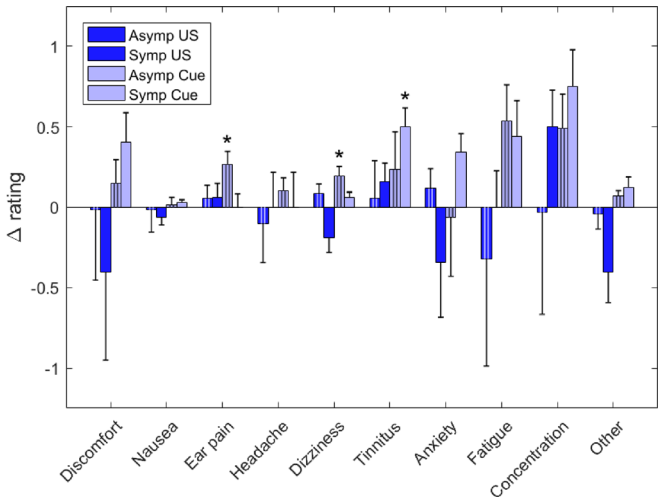

FIG. 3. (Color online) Average difference in baseline-corrected ratings for conditions A and B and X and Y averaged over all trials for the symptomatic (eight participants) and asymptomatic (32 participants) groups. For the cue conditions, a positive difference means that the "Ultrasound ON" cue yielded a higher rating than the "Ultrasound OFF" cue (i.e., that a nocebo effect was present). Statistically significant differences ($p < 0.05$) are marked with an asterisk. Error bars show the standard error of the mean across participants.

either the asymptomatic ($p = 0.12$) or symptomatic ($p = 0.27$) group.

Small but statistically significant effects of the false "Ultrasound ON" cue condition compared to the true "Ultrasound OFF" cue condition were found for ear pain ($Q(1) = 9.0$, $p = 0.004$, mean effect: 0.27 rating points) and dizziness [$Q(1) = 9.0$, $p = 0.004$, mean effect: 0.2 rating points] in the asymptomatic group and for tinnitus [$Q(1) = 5.0$, $p = 0.025$, mean effect: 0.5 rating points] in the symptomatic group. After completing the experiment, five asymptomatic participants and one symptomatic participant revealed to the experimenter that they were suspicious that US had either not been presented in the false-cue condition or had been presented in the true-cue condition. Conversely, some participants reported a high level of certainty that US had been presented in the false-cue condition. Some of these participants reported additional effects to the experimenter at the end of the true-cue sessions, despite giving zero ratings for "other symptoms" within that condition. One symptomatic participant reported tightness in the chest, shortness of breath, increased heart rate, and headache. Two asymptomatic participants reported increased itchiness throughout the session, one tightness in the chest, and one reported feeling "like gravity had shifted."

2. Performance on sustained attention task

No effect of US on either response times or commission errors was found for the SART in either group using a Friedman's analysis (Fig. 4). Uncorrected Spearman correlation analyses suggested a correlation between differences in anxiety rating between the US exposure and sham conditions (A and B) and differences in SART response times between the same conditions ($r_s = 0.57$; $p = 0.0007$) for the asymptomatic group. Evidence of a correlation was also found between the differences in SART commission errors and differences in concentration ratings for the cued conditions (X and Y; $r_s = 0.37$; $p = 0.0007$), for the asymptomatic group.

3. Galvanic skin responses

A repeated-measured analysis of variance (ANOVA) suggested that the baseline-corrected GSR estimates were significantly elevated by the false-cue for the asymptomatic group ($F(1,30) = 4.46$, $p = 0.004$), but not the symptomatic group ($F(1,6) = 1.14$, $p = 0.328$). No effects of US were found for either group, either when looking at the mean effect over each trial (see Fig. 5), or when looking at the time histories within each trial. For the asymptomatic group, evidence of a negative correlation was found between the differences in ratings of fatigue for the US conditions and the differences in GSR for the US conditions ($r_s = -0.51$; $p = 0.0027$). This negative correlation may be expected as skin conductance has been shown to decrease with physiological relaxation, in periods of rest or sleep (Malmo, 1959).

C. Case study

1. Subjective ratings

The results of a case study on a participant who reported particularly significant and consistent effects of US are shown in Fig. 6. No evidence of an effect of US, which was presented at an SPL of 94 dB in 1.5 min repeated trials, was found. Some evidence of small nocebo effects was found for some ratings. The participant was unable to identify the condition that contained US above chance levels (33.3% correct; one-sample Wilcoxon signed-rank test, $p = 0.2$). No clear trends across conditions were observed for the SART attention measure or the GSR.

IV. FURTHER DISCUSSION

The severe effects that have been reported by some members of the public were not reproduced in this study, either after genuine US exposure under double-blind conditions, or as a result of a false cue, which informed the participant that US was present when it was not. There are several possible reasons why an effect was not found. A first possibility is that the SPL of the stimulus in the current experiment was not sufficiently high, the stimulus duration was not sufficiently long, or the effect does not occur, or is very small, for a tone at 20 kHz. A second possibility is that anecdotal reports from the public come from a particularly sensitive minority, and that they were not represented in our sample (or were a minority whose responses were diluted through averaging). This possibility was reduced, but not eliminated, by selecting a group that had self-reported as sensitive, and by adding an additional case study of an individual who reported being particularly sensitive. Previous studies have not taken such measures [Leighton (2016), for example, has criticized past studies, such as Knight (1968), for failing to take these measures]. A final possibility is that the anecdotally reported effects do not result from exposure to US.

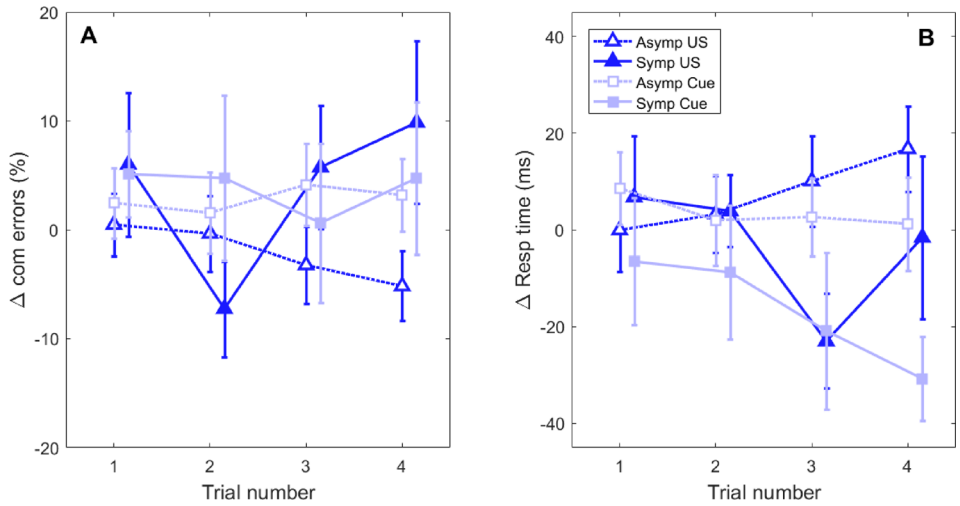

FIG. 4. (Color online) Mean change in performance metrics for the SART, averaged over participants for the two groups (32 in asymptomatic group; 8 in symptomatic group). Panel A shows the change in the percentage of commission errors between the US exposure and sham conditions (darker lines and symbols) and the false and true cue conditions (lighter lines and symbols). Panel B shows the change in reaction time between these conditions. A positive value indicates that the response time or percentage of commission errors was highest in the US exposure or false-cue condition. Markers are horizontally offset for clarity. Error bars show the standard error of the mean across participants.

When considering the first possibility, it should be noted that some of the reports received by this research group have been of symptoms produced on a much shorter time scale (minutes or even seconds) by sources producing tones at 20 kHz that would be expected to have a significantly lower SPL than those used in the present study. One example of such a source is PAVA systems, which commonly use 20 kHz tones at SPLs between 40 and 80 dB (most often below 55 dB) to monitor their operational status (Mapp, 2016, 2017; Fletcher *et al.*, 2018b). However, for other cases of symptoms that may have been produced by exposure to inaudible US [particularly in occupational settings, where exposures to SPLs of up to 133 dB at 20 and 25 kHz have been reported (Skillern, 1965) or in close proximity to pest deterrents (Leighton, 2016)]; it remains plausible that the SPL and exposure duration used in the current study (which were restricted in consideration of participant safety) were not sufficient to produce symptoms. Other differences between the exposure over headphones in the current study, and that experienced in real-world settings, are the effects caused by room acoustics and the individual's head-related transfer function on the sound level that reaches the ears [which would be expected to change as the individual moves around the environment, an effect that could show greater unpredictability at very high frequencies (Leighton, 2016)].

The second of the possibilities listed above is that genuine effects are produced as a result of US exposure at the SPL, frequency and duration used in the current experiment, but that the current study did not have a sufficient sample of participants who genuinely experienced these effects. Given the difficulties in recruiting participants (partly due to the

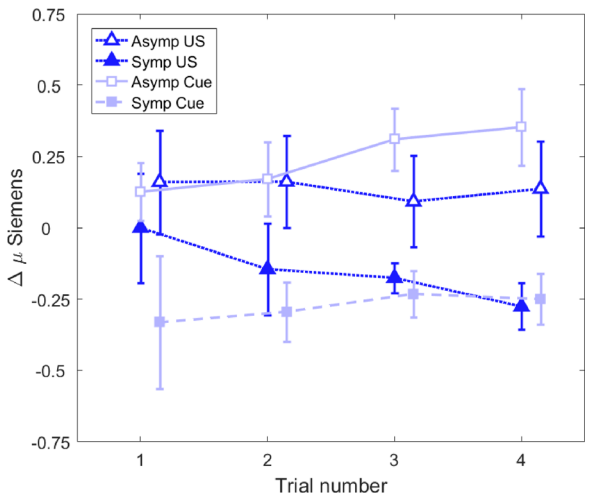

FIG. 5. (Color online) Mean difference in galvanic skin response between the two US conditions (A and B; darker lines and symbols) and cue conditions (X and Y; lighter lines and symbols) for each trial for the asymptomatic (32 participants) and symptomatic (10 participants) groups. A positive change means that the galvanic skin response was largest in the US or false-cue condition. Markers are horizontally offset for clarity. Error bars show the standard error of the mean across participants.

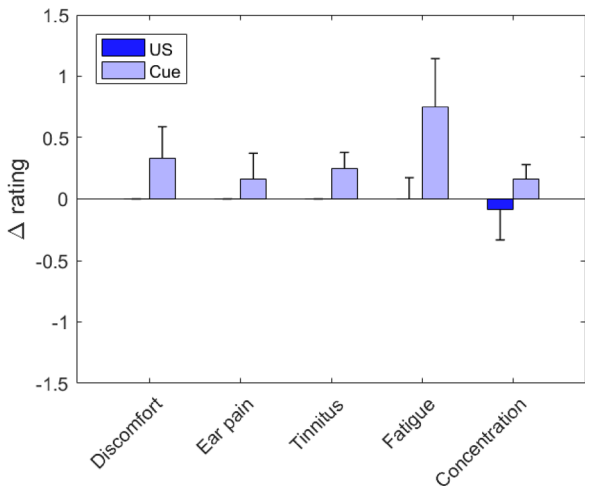

FIG. 6. (Color online) Average difference in ratings for the two US conditions (darker bars) and two cue conditions (lighter bars). All of the symptom ratings for the case study that were not zero on average across trials are shown. For the cue conditions, a positive effect means that the "Ultrasound ON" cue produced a larger effect than the "Ultrasound OFF" cue. Error bars show the standard error of the mean across all trials.

need for multiple sessions on separate days and the lack of people locally who specifically report symptoms from US exposure), it is not possible conclusively to rule out this possibility. A future study across multiple centers might be able to overcome this recruitment difficulty.

The final possible explanation of the effects measured is that no genuine symptoms resulting from exposure to inaudible US exist, and that the false-cue was not a sufficiently strong stimulus to produce a nocebo effect as strong as that produced in real-world settings. One reason for a reduced nocebo effect might be significant levels of skepticism surrounding the trustworthiness of the false and/or true cue within the participant sample. Informal reports to the experimenter from some participants of suspicion of the cues supports this possibility. A further alternative is that strong nocebo effects by US devices may be more readily produced either outside of a controlled laboratory setting, or when a stronger nocebo cue is used. Such stronger cues may include a visible dummy device that has previously been reported to produce symptoms, or the explicit statement that exposure to US might cause harmful effects (see, for example, Witthoft and Rubin, 2013).

It should be noted that the symptoms reported for US sources, including headache, fatigue, dizziness, nausea, skin itching or tingling, and concentration difficulties, have also been reported as a result of exposure to radio frequencies, mobile phones, WIFI, and a range of other wireless devices, all of which have been repeatedly found not to produce effects under double-blind conditions, with symptoms often being found under sham conditions (for reviews, see Oftedal *et al.*, 2000; Koivisto *et al.*, 2001; Rubin *et al.*, 2006; Roosli, 2008; Rubin *et al.*, 2010). Furthermore, there has been a significant body of work on risk perception, which has resulted in the identification of the characteristics of a new technology that make it likely to cause public concern (Slovic, 1987). These include: the perception that people exposed to the technology do not know they are being exposed (because exposure is invisible), exposure being involuntary, the perceived risk being unknown to science, the perceived risk affecting children, and the exposure being perceived to be capable of causing delayed or hidden health effects. These perceived characteristics apply to reports we have received of exposure from inaudible US. It remains possible that the reported effects are the result of a misattribution of symptoms.

While this study found no evidence of effects of exposure to inaudible US, it should be understood that a stimulus SPL of 84 dB (the average level of the sound presented in this study) at 20 kHz would be audible for a significant proportion of, particularly young, individuals (Henry and Fast, 1984; Ashihara, 2006; Ashihara *et al.*, 2006; Ashihara, 2007; Rodriguez Valiente *et al.*, 2014). Thus, subjective effects may be produced in some individuals at the frequency and average exposure level used in this study.

Future studies must address several practical issues encountered in this study. One is the need to recruit greater numbers of participants, and in particular to test the most sensitive subjects (Leighton, 2016). This is common to studies over many decades of human adverse effects to ultrasonic exposure in air. One way would be to design a future study with a travelling acoustic test booth, since some participants who self-reported as sensitive declined to participate because of fears of exposure on the journey. Another is to recognize the fact that public exposures are at times exceeding the maximum permitted levels (MPLs; Leighton, 2016), yet ethically we cannot collect data under conditions which exceed (in terms of intensity and/or duration) MPLs (and particularly, cannot collect data for children, a cohort of particular concern). It is unethical not to address the conundrum that we cannot ethically collect safety data for exposures that are occurring in public places.

ACKNOWLEDGMENTS

This research was primarily funded by the Colt Foundation (Ref. No. CF/03/15). Additional financial support from the European Metrology Programme of Innovation and Research (EMPIR, Grant No. 15HLT03 EARII) for Timothy Leighton's and Ben Lineton's time is gratefully acknowledged. The EMPIR programme is co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme. The datasets generated and analysed during the current study are openly available from the University of Southampton repository at http://dx.doi.org/10.5258/SOTON/.

APPENDIX A: PARTICIPANT SCREENING AND GROUP ALLOCATION

Please fill in the following questionnaire to determine your eligibility for this experiment. If yes to any of the questions 2 to 14, please give additional details below. All data will be kept confidential.

1. What is your age in years? _____ years

2. Do you have a hearing impairment that you are aware of? **Yes / No**

3. Do you have, or have you recently had any pain, tenderness, infections, discharge, surgery or bleeding in either of your ears? **Yes / No**

4. Do you have a history of frequent exposure to loud noise? **Yes / No**

5. Do you take any ototoxic medications (e.g., aminoglycoside antibiotics, such as gentamicin)? **Yes / No**

6. Do you experience tinnitus (ringing, buzzing, whistling or any other sounds in either of your ears)? **Yes / No**

7. Do you suffer from hyperacusis (reduced tolerance and increased sensitivity to everyday sounds)? **Yes / No**

8. Have you been exposed to loud sounds in the past 24 hours? **Yes / No**

9. Do you expect to be exposed to loud sounds in the next 24 hours (e.g., visiting a night club, or concert, or taking part in an experiment involving high levels of sound presentation?) **Yes / No**

10. Do you suffer from epilepsy? **Yes / No**

11. Have you ingested a significant amount of caffeine in the last two hours, e.g., drank more than one cup of tea or coffee, drunk an energy drink, or taken pro plus)? **Yes / No**

12. Have you taken recreational drugs in the last week? **Yes / No**

13. Have you drunk more than 6 units of alcohol (more than 2 pints of beer or 2 standard glasses of wine) in the last 24 hours? **Yes / No**

14. Have you undertaken strenuous physical or mental activity today? **Yes / No**

If you have answered "yes" to any of questions 2–14, please give further details below.

Details for question number (s): _____:

15. Have you ever experienced unpleasant symptoms that you believe were caused by exposure to very high-frequency sound? **Yes / No**

If you have answered "yes" to question 15, please give further details below. If possible, include answers to the following:

a) What is the nature of these symptoms?

b) How long ago, approximately, did you first experience them?

c) In general, do/did the symptoms arise as soon the exposure began, or only after a period of time?

d) In general, how long did/do the symptoms endure after exposure has ceased?

e) What type of device/devices do you suspect have caused the symptoms, if known? (e.g., pest scarers).

Details (if "Yes"):

16. Do you believe that your hearing abilities at very-high frequencies is particularly good (e.g., do you believe you can hear sounds at high frequencies that most people cannot)?

Yes/No

17. Have you ever experienced unpleasant symptoms that you believe were caused by exposure to ultrasonic devices (devices producing sounds too high in frequency for you to hear)? **Yes / No**

If you have answered "yes" to question 17, please give further details below. If possible, include answers to the following:

f) What is the nature of these symptoms?

g) How long ago, approximately, did you first experience them?

h) In general, do/did the symptoms arise as soon the exposure began, or only after a period of time?

i) In general, how long did/do the symptoms endure after exposure has ceased?

j) What type of device/devices do you suspect have caused the symptoms, if known? (e.g., pest scarers).

Details (if "Yes"):

18. Do you have any expectation of symptoms that you might experience during testing with sounds that are at ultrasonic frequencies **Yes / No**

Details (if "Yes"):

APPENDIX B: SUBJECTIVE RATING SCREENING QUESTIONNAIRE

Please rate your overall discomfort level										
0	1	2	3	4	5	6	7	8	9	10
None										*Severe*
Over the last 4 minutes I experienced[a]**. . .**										
Nausea										
0	1	2	3	4	5	6	7	8	9	10
Not at all										*Severe*
Pain, pressure, or fullness in one or both ears										
0	1	2	3	4	5	6	7	8	9	10
Not at all										*Severe*
Headache/ pain or pressure somewhere other than my ears										
0	1	2	3	4	5	6	7	8	9	10
Not at all										*Severe*
Dizziness or light-headedness										
0	1	2	3	4	5	6	7	8	9	10
Not at all										*Severe*
Tinnitus (ringing, buzzing, or other sounds in my ears)										
0	1	2	3	4	5	6	7	8	9	10
Not at all										*Severe*
Anxiety										
0	1	2	3	4	5	6	7	8	9	10
Not at all										*Severe*
Fatigue										
0	1	2	3	4	5	6	7	8	9	10
Not at all										*Severe*
Other symptoms										
0	1	2	3	4	5	6	7	8	9	10
Not at all										*Severe*

[a]This phrase was changed to "over the last hour I experienced…" for the screen phase.

[1]In this paper, the term "ultrasound" is taken to mean frequencies in excess of 17.8 kHz, following Leighton (2017) who argued that ultrasonic regulations to date extended down to the lower frequency limit of the third octave band centered in 20 kHz.

Acton, W. I. (**1974**). "The effects of industrial airborne ultrasound on humans," Ultrasonics **12**, 124–128.

Acton, W. I. (**1983**). "Exposure to industrial ultrasound: Hazards, appraisal and control," J. Soc. Occup. Med. **33**, 107–113.

Acton, W. I., and Carson, M. B. (**1967**). "Auditory and subjective effects of airborne noise from industrial ultrasonic sources," Br. J. Ind. Med. **24**, 297–304.

Ashihara, K. (**2006**). "Threshold of hearing for pure tones between 16 and 30 kHz," J. Acoust. Soc. Am. **120**, 3245.

Ashihara, K. (**2007**). "Hearing thresholds for pure tones above 16 kHz," J. Acoust. Soc. Am. **122**, EL52.

Ashihara, K., Kurakata, K., Mizunami, T., and Matsushita, K. (**2006**). "Hearing threshold for pure tones above 20 kHz," Acoust. Sci. Technol. **27**, 12–19.
British Society of Audiology (**2017**). *Recommended Procedure: Pure-tone Air-conduction and Bone-conduction Threshold Audiometry with and without Masking* (British Society of Audiology, Reading, UK).
Crabtree, R. B., and Forshaw, S. E. (**1977**). "Exposure to ultrasonic cleaner noise in the Canadian forces," DCIEM technical report no. 77X45.
Ebelthite, H. (**2016**). "Could your health be ruined by noises you can't hear? Some gadgets emit silent ultra-high whines that may hurt you," The Daily Mail. Available from: http://www.dailymail.co.uk/femail/article-3527060/Could-health-ruined-noises-t-hear-gadgets-emit-silent-ultra-high-whines-hurt-you.html [Last viewed 2 August 2017].
Fletcher, M. D. (**2016**). "Effects of ultrasound and very high-frequency sound," Facebook (public group). Available from: https://www.facebook.com/groups/249882768679771/ (Last viewed 29 July 2017).
Fletcher, M. D., Lloyd Jones, S., White, P. R., Dolder, C. N., Lineton, B., and Leighton, T. G. (**2018a**). "Public exposure to ultrasound and very high-frequency sound in air," J. Acoust. Soc. Am. **144**(4), 2554–2564.
Fletcher, M. D., Lloyd Jones, S., White, P. R., Dolder, C. N., Leighton, T. G., and Lineton, B. (**2018b**). "Effects of very high-frequency sound and ultrasound on humans. Part I: Adverse symptoms after exposure to audible very high-frequency sound," J. Acoust. Soc. Am. **144**(4), 2511–2520.
Foxe, J. J., Morie, K. P., Laud, P. J., Rowson, M. J., de Bruin, E. A., and Kelly, S. P. (**2012**). "Assessing the effects of caffeine and theanine on the maintenance of vigilance during a sustained attention task," Neuropharmacology **62**, 2320–2327.
Hardin, J. W., and Hilbe, J. M. (**2003**). *Generalized Estimating Equations* (Chapman and Hall, New York).
Henry, K. R., and Fast, G. A. (**1984**). "Ultrahigh-frequency auditory thresholds in young adults: Reliable responses up to 24 kHz with a quasi-free field technique," Audiology **23**, 477–489.
Herman, B. A., and Powell, D. (**1981**). "Airborne ultrasound: Measurement and possible adverse effects," U. S. Department of Health and Human Resources, Public Health Servicen Publication, Rockville, MD (Food and Drug Administration).
Knight, J. J. (**1968**). "Effects of airborne ultrasound on man," Ultrasonics **6**, 39–41.
Koivisto, M., Haarala, C., Krause, C. M., Revonsuo, A., Laine, M., and Hamalainen, H. (**2001**). "GSM phone signal does not produce subjective symptoms," Bioelectromagnetics **22**, 212–215.
Kucera, P., Goldenberg, Z., and Kurca, E. (**2004**). "Sympathetic skin response: Review of the method and its clinical use," Bratisl. Lek. Listy **105**, 108–116.
Lawton, B. W. (**2001**). "Damage to human hearing by airborne sound of very high frequency or ultrasonic frequency," Health and Safety Executive Contract Research Report No. 343/2001.
Leighton, T. G. (**2007**). "What is ultrasound?," Progr. Biophys. Mol. Biol. **93**, 3–83.
Leighton, T. G. (**2016**). "Are some people suffering as a result of increasing mass exposure of the public to ultrasound in air?," Proc. R. Soc. A **472**(2185), 20150624.
Leighton, T. G. (**2017**). "Comment on 'Are some people suffering as a result of increasing mass exposure of the public to ultrasound in air?'," Proc. R. Soc. A **473**(2199), 20160828.
Levitt, H. (**1971**). "Transformed up–down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, Suppl. 2, 467.
Maccà, I., Scapellato, M. L., Carrieri, M., Maso, S., Trevisan, A., and Bartolucci, G. B. (**2015**). "High-frequency hearing thresholds: Effects of age, occupational ultrasound and noise exposure," Int. Arch. Occup. Environ. Health **88**, 197–211.
Malmo, R. B. (**1959**). "Activation: A neurophysiological dimension," Psychol. Rev. **66**, 367–386.
Manly, T., Lewis, G. H., Robertson, I. H., Watson, P. C., and Datta, A. (**2002**). "Coffee in the cornflakes: Time-of-day as a modulator of executive response control," Neuropsychologia **40**, 1–6.
Manly, T., Owen, A. M., McAvinue, L., Datta, A., Lewis, G. H., Scott, S. K., Rorden, C., Pickard, J., and Robertson, I. H. (**2003**). "Enhancing the sensitivity of a sustained attention task to frontal damage: Convergent clinical and functional imaging evidence," Neurocase **9**, 340–349.
Manly, T., Robertson, I. H., Galloway, M., and Hawkins, K. (**1999**). "The absent mind: Further investigations of sustained attention to response," Neuropsychologia **37**, 661–670.
Mapp, P. (**2016**). "Ultrasonic surveillance monitoring of PA systems, a safety feature or audible hazard?," Proc. Inst. Acoust. **38**(Part 2), 1–16.
Mapp, P. (**2017**). "Potential audibility and effects of ultrasonic surveillance monitoring of PA and life safety sound systems," in *Audio Engineering Society Convention Paper 9776 Presented at the 142nd Convention 2017* (Berlin, Germany).
Oftedal, G., Wilen, J., Sandstrom, M., and Mild, K. H. (**2000**). "Symptoms experienced in connection with mobile phone use," Occup. Med. (London) **50**, 237–245.
Pharris, C. (**1948**). "Ultrasonic sickness," Am. Ind. Hyg. Assoc., Q. **9**(3), 57–62.
Robertson, I. H., Manly, T., Beschin, N., Daini, R., Haeske-Dewick, H., Homberg, V., Jehkonen, M., Pizzamiglio, G., Shiel, A., and Weber, E. (**1997**). "Auditory sustained attention is a marker of unilateral spatial neglect," Neuropsychologia **35**, 1527–1532.
Rodriguez Valiente, A., Trinidad, A., Garcia Berrocal, J. R., Gorriz, C., and Ramirez Camacho, R. (**2014**). "Extended high-frequency (9–20 kHz) audiometry reference thresholds in 645 healthy subjects," Int. J. Audiol. **53**, 531–545.
Roosli, M. (**2008**). "Radiofrequency electromagnetic field exposure and non-specific symptoms of ill health: A systematic review," Environ. Res. **107**, 277–287.
Rubin, G. J., Hahn, G., Everitt, B. S., Cleare, A. J., and Wessely, S. (**2006**). "Are some people sensitive to mobile phone signals? Within participants double blind randomised provocation study," BMJ [Br. Med. J.] **332**, 886–891.
Rubin, G. J., Nieto-Hernandez, R., and Wessely, S. (**2010**). "Idiopathic environmental intolerance attributed to electromagnetic fields (formerly 'electromagnetic hypersensitivity'): An updated systematic review of provocation studies," Bioelectromagnetics **31**, 1–11.
Skillern, C. P. (**1965**). "Human response to measured sound pressure levels from ultrasonic devices," Am. Ind. Hyg. Assoc. J. **26**, 132–136.
Slovic, P. (**1987**). "Perception of risk," Science **236**, 280–285.
Smilek, D., Carriere, J. S., and Cheyne, J. A. (**2010**). "Failures of sustained attention in life, lab, and brain: Ecological validity of the SART," Neuropsychologia **48**, 2564–2570.
Stelmachowicz, P. G., Beauchaine, K. A., Kalberer, A., Kelly, W. J., and Jesteadt, W. (**1989**). "High-frequency audiometry: Test reliability and procedural considerations," J. Acoust. Soc. Am. **85**, 879–887.
Ueda, M., Ota, A., and Takahashi, H. (**2014**). "Investigation on high-frequency noise in public space.—We tried noise abatement measures for displeasure people," in *Proc. of the 7th Forum Acusticum*, Krakow, Poland, 7–12 September 2014.
Witthoft, M., and Rubin, G. J. (**2013**). "Are media warnings about the adverse health effects of modern life self-fulfilling? An experimental study on idiopathic environmental intolerance attributed to electromagnetic fields (IEI-EMF)," J. Psychosom. Res. **74**, 206–212.