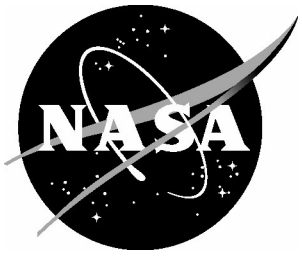
NASA/TM–20220012593



Consistency of Annoyance Perception of Sonic Booms and Survey Satisficing

Anvitha Chintam
University of Maryland, College Park, Maryland

Kathryn M. Ballard, William J. Doebler, and Aaron B. Vaughn
Langley Research Center, Hampton, Virginia

December 2022

NASA STI Program Report Series

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

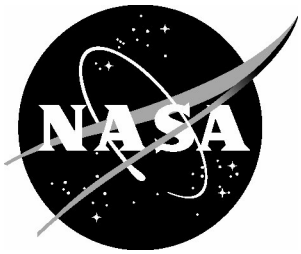
For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- Help desk contact information:

https://www.sti.nasa.gov/sti-contact-form/
and select the "General" help request type.

NASA/TM–20220012593



Consistency of Annoyance Perception of Sonic Booms and Survey Satisficing

Anvitha Chintam
University of Maryland, College Park, Maryland

Kathryn M. Ballard, William J. Doebler, and Aaron B. Vaughn
Langley Research Center, Hampton, Virginia

National Aeronautics and Space Administration

Langley Research Center Hampton,
Virginia 23681-2199

December 2022

Acknowledgments

This work was supported by NASA's Commercial Supersonic Technology Project. The authors are grateful for the input of the Commercial Test Planning and Execution Team Members including Nathan Cruze and Jonathan Rathsam, as well as Prof. Frederick Conrad of the University of Maryland. The authors also thank Ran Cabell and Steve Bauer for their review and helpful suggestions.

The use of trademarks or names of manufacturers in this report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

Table of Contents

Abstract

There is a potential in community sonic boom tests for survey measurement error due to recall errors. This occurs when a reported value differs from a participant's true experience. The two topics examined in this paper are 1) recall bias: how consistently participants recall their annoyance to sonic thump events and 2) survey satisficing: whether participants opt for early survey termination if the option is available. Data from QSF18 were examined for evidence of recall bias and survey satisficing via inconsistencies between single event and daily summary survey reports. In terms of recall bias, when only one single event survey and daily summary survey were submitted by a participant on a particular day, both the single event and daily summary annoyance ratings match in the majority of instances (167 of 186). In terms of survey satisficing, there were fewer questions in the daily summary survey if the participant reported not hearing any sonic thumps during the day. Nevertheless, the instances of participants inconsistently reporting boom audibility between the single event and daily summary survey are relatively few (66 of 767). Therefore, the results of this study provide evidence to rule out recall bias and satisficing as sources of error in the QSF18 study.

Introduction

Although supersonic flight offers much shorter travel times, overland commercial supersonic flight was prohibited by the Federal Aviation Administration (FAA) in 1973 due to concerns regarding the loud sonic boom produced by these aircraft. A sonic boom results when an object travels supersonically and pressure waves combine together forming a shock wave, which results in a loud boom when the shock wave reaches the ground (Rathsam and Wilson, 2019). In the past few years, NASA has been developing the X-59 aircraft, which utilizes shaped-boom technology to decrease the loudness of these pressure waves, thereby reducing the sonic boom to a milder "sonic thump" when it reaches the ground.

Empirical evidence is crucial for informing potential future overland commercial supersonic aircraft certification standards. NASA is slated to gather community feedback data from 2024 to 2026 as the X-59 aircraft flies over designated communities in the contiguous United States. This requires accurate survey methods to measure public perception of the quieter sonic thump. NASA previously utilized F-18 dive maneuvers (Haering et al. 2006) to simulate sonic thumps during the Quiet Supersonic Flights 2018 (QSF18) pilot study, which involved gathering survey data that asked participants how annoyed they were by the sonic thumps. Participants were asked to complete "single event" surveys following each sonic thump that occurred during the day as well as a "daily summary" survey at the end of each day describing their overall perception of the sonic thumps. The respondents were exposed to up to eight sonic thumps per day over eleven days. Analysis of the QSF18 respondent experience may provide insights for the survey instruments deployed during the X-59 community studies. Certain survey administration conditions can cause measurement error and ultimately result in an inaccurate measure of the construct. A preliminary literature review provides the basis for prominent error types that can produce bias in survey data. This work addresses two contributors of measurement error: recall error[1] and survey satisficing.[2] The first pertains to how consistently participants score annoyance between the single event and daily summary surveys. The latter is more specific to how survey satisficing plays a role in how participants respond to the questionnaires. The data is more likely biased when participants have recall errors or frequently satisfice. In this case, biased data may mean that annoyance is not accurately measured due to forgetting or misreporting by the sampled population and the perception of booms by the general public may differ from what was initially reported in QSF18 studies. This paper addresses these measurement errors through two primary research questions:

1) How consistently do survey participants recall their annoyance to sonic thumps over the course of a few hours?

2) Will participants opt for a survey response that allows for early termination if the option to do so exists?

[1] Recall error is the inability of respondents' to accurately remember the variable of interest.
[2] "Survey satisficing" is defined as the tendency to provide satisfactory but not optimal responses to reduce their effort (Zhang & Conrad, 2014; Krosnick 1991).

Objective 1: Recall Bias

The first research question focuses on how consistently survey participants recall their perception of annoyance to sonic booms over the course of a few hours. This is of interest because circumstances (e.g., work meeting, phone call, driving, etc.) may prevent a prompt response from a respondent to a single event survey. These respondents may respond later, and their delayed responses are used to determine the validity of their recollection of their annoyance.

Literature Review

Survey researchers have widely recognized recall bias as being a common source of measurement error. Recall bias is a systematic error that arises when participants do not remember past experiences or events or omit information. Subsequent events can greatly impact how a person encodes,[3] processes, and retrieves information for survey purposes. This is of interest for the X-59 community tests as they will be longitudinal and rely on the participants' ability to recall whether a boom disturbed them in any way.

Literature on recall bias suggests an inverse relationship between the length of time provided for survey responses and the accuracy of reports collected (Clarke et al. 2008). More specifically, longer recall periods[4] are more prone to cause recall bias. The Survey of Living Conditions conducted by Statistics Sweden collected survey data to address recall bias. The data consist of patients' ability to recall annual drug utilization and hospitalization and were used to create a statistical framework that would aid in identifying a maximum recall period for varying reference periods[5]: an hour, 24 hours, and a week. They report on a trade-off between a shorter recall window that may result in fewer responses and a longer window that can introduce recall bias. Though the current study consisted of a longer time frame than of X-59 community surveys, this study highlights the importance of choosing a recall window that is optimal in order to keep error and bias low while collecting a sufficient number of responses.

Another potential concern for recall bias in X-59 community surveys is background noise in relation to response latency. Background noise is a particular concern for X-59 community tests as it may lead to issues with identifying or paying attention to a single boom event. Researchers in a study evaluated perception of an auditory stimulus with background noise (Davenport, 1972). The study set out to investigate how different types of background noise affect detection rates of a signal. Subjects responded to a signal and were told to press a response key as quickly as possible when they heard the signal. Participants were tested under four conditions of background noise: silence, broadband noise, continuous music, and randomized music. Results showed that despite the changes in background noise, respondents were able to detect the signal accurately and had relatively low response latencies – meaning that they pressed the key promptly after hearing a signal. In the study, participants were informed in advance of the signal, which may also contribute to quick responses. While notifying

[3] Encoding is a psychological concept involving the "conversion of sensory input into a form where it can be processed and deposited in memory" (American Psychological Association).
[4] Recall period is the amount of time elapsed between the reference period and data collection.
[5] The reference period is the time frame for which survey respondents are asked to report activities or experiences for the variable of interest (Clarke et al., 2008).

participants ahead of time of sonic booms in a community survey may reduce response latency, it is not a viable option as it would introduce bias compared to a uniformed population.

Perception of annoyance is a highly subjective measurement and differs greatly from the task of hearing a signal and reporting audibility. This is especially true since noise annoyance can lead to an emotional response such as irritability due to disruption. Research on this topic proposes that emotional arousal and verbal context affect memory enhancement in information processing (Medford et al., 2007). This was measured in a study by having participants read a selection of sentences that were separated into emotional and neutral conditions. One group was given a sentence containing an emotional "target" word, which was meant to be emotionally arousing, while the other would have a "neutral" word in place of the target word that did not alter the meaning of the overall sentence. Participants were then put into three groups and surprised with memory tasks, after either 5 minutes, an hour, or 24 hours. Participants' recall was evaluated in two assessments: first, they were presented with words and asked to indicate the ones they had previously seen, known as a recognition task. In the second, they were asked to freely recall[6] the words. In either condition, target words were remembered significantly better despite differences in the time elapsed. Recognition of emotional words was stable regardless of time delay group. However, free recall data indicated that emotional enhancement was significant at 24 hours but not after an hour.

The findings of the Medford et al. (2007) study indicate emotional enhancement is important to information processing and can influence both the memory and the emotional response to get stronger over time. Thus, the working hypothesis is that sonic boom survey participants who respond promptly to a single event boom may not have had time to process their emotional response since submission is encouraged as soon as possible, with responses on the order of a few minutes to hours after they heard the thump. However, the ramifications of this may be that if someone found the boom annoying, the memory of that noise event in an emotional context will get stronger after several hours rather than one hour. As a result, the most pressing concern is that participants may misreport annoyance if the boom was encoded in an emotional context such that the time required to realize their true annoyance would be after submitting the single event survey. It is possible to test this to some extent in sonic boom community survey data by searching for increased annoyance reported in the daily summary survey compared to a single event survey. Although events encoded with an emotional component add another facet to retrieval of a single event, annoyance is still not itself defined as emotion and more so relies on subsequent events to bring about an emotional context. This leads to further questioning of whether encoding a sonic boom in an emotional context is a precondition for acoustic salience[7].

A study regarding acoustic salience addresses the theory that animals have evolved to make emotionally intense vocalizations in order to eliminate the possibility of listener habituation (Anikin, 2020). Researchers tested 128 human non-verbal vocalizations representing eight emotions: amusement, anger, disgust, effort, fear, pain, pleasure, and sadness. They not only

---

[6] Free recall is a psychological concept defined as a type of memory task in which participants attempt to remember previously studied items in any order (American Psychological Association).
[7] Salience is "the capacity for sensory stimuli to command attention in bottom-up, involuntary fashion" (Anikin, 2020).

looked at emotional context affecting salience, but they also measured salience independently. Anikin used two independent samples of participants to provide measures of salience. These measures were self-reported and objective salience, the latter was operationalized as the reduction in performance in a short-term memory task. The two measures were then compared to ratings of emotional intensity and acoustic characteristics of the different stimuli.

Participants were asked to memorize and repeat a sequence of six numbers they heard through earphones in one channel, while simultaneously being told to ignore vocalizations in the other. The sequence was vocalized with unpredictable timing and could play in either the left or right ear. The distractor vocalizations were played concurrently with background noise in both channels. Though there would be overlap with the distractor vocalizations and the target sequence, they were never played within the same channel. Results showed that self-reported high salience produced 25% more recall errors in the short-term memory task, but emotional intensity had no independent effect on recall errors. Participants were able to recognize the target sequences and recall them with high accuracy – 4.8 out of 6 digits on average, and all six were recalled correctly in over half of the trials.

Researchers concluded that emotionally intense vocalizations were more effective at attracting the listeners' attention but were not the reason the sound was salient. Other factors influencing the recall error and salience were acoustic characteristics of vocalizations such as greater duration and intensity, higher pitch, bright timbre, and rapid modulations of sound. The author also proposed that salience is more closely intertwined with the modulations of a sound, which suggests that noise events would have correct latencies despite not being encoded in an emotional context (Anikin, 2020).

Implications for the sonic boom community tests stem from the recall error identified. Anikin concluded that high-salience sounds produced 25% more recall error in a short-term memory task. This raises a concern with community test surveys as participants will be prompted to respond as soon as possible after a sonic thump. Thus, accurate recall of this noise event may have been an issue in the past and may be a concern in the future surveys. However, respondents in the study were given multiple stimuli and asked to recall them, whereas the sonic boom community testing would only focus on one noise event at a time and recall might be easier for a single noise event rather than six.

To investigate other factors that affect the relationship between memory and response latency, researchers hypothesized that performance on a test of memory utilizes a unidimensional structure (MacLeod and Nelson, 1984). They suggested that if strength of memory increases for a set of items, then the probability of error on those items should decrease and vice versa. The goal of their research was to address this assumption by investigating the relationship between response accuracy and response latency. They used three conditions with a variety of memory tasks given to the participants within each experiment. In the first experiment, the independent variable was the length of long-term retention that followed learning. In the second experiment, the independent variable was the type of processing (semantic vs. non-semantic) during learning. In the final experiment, the independent variable was the number of study trials versus test trials during learning.

In the first experiment, 74 students participated in a memory task where they were asked to remember number-noun pairs. Researchers evaluated the accuracy-latency relationship with

different recall periods (1 week, 3 weeks, or 5 weeks). In the second experiment, 27 students participated in a task where they were shown two words and asked if the words had the same number of syllables. They had to respond yes or no as quickly as possible. In the final experiment, 58 students were split into two groups and participated in a memory task where they were given 16 items to remember and had 5 minutes to state which noun went with which number stimulus – similar to the first experiment.

Results of all three experiments showed a decrease in correct latency, i.e., the time it takes for a respondent to respond correctly, across all test groups and error latencies, i.e., the time it takes for a respondent to respond incorrectly, were significantly longer than corresponding correct latencies. Error probability was lower for the multiple-study group in the third experiment, meaning the more learning participants did the less the error probability. They also reported that accuracy and correct latency are not unidimensional.

The authors suggest that response latencies should be separated as correct latency and error latency. Previously, it was assumed that a combined latency of the two increased as error probability increased; however, this is not entirely accurate. This study found that error latency alone impacts error probability and that the combined latency reflects this trend since it includes the error latency. They also mention that error latency and correct latency reflect different facets of memory and tap different dimensions of the memory structure. Encoding is measured by error latency, while correct latency measures the number of decoding steps during retrieval before an item is output. Therefore, they proposed that time elapsed can increase error latency, while accuracy and correct response latency do not influence each other. This work on response times and accuracy suggests that when respondents are pressured to respond quickly to a question, they are more prone to recall errors. Community test surveys should be administered with care taken to guard against this type of error, which may mean that speed of responses should not be prioritized.

Recall bias occurs in surveys when participants do not recall past experiences and may not report their response accurately. The literature suggests that longer recall periods, the time between the event of interest and the time data is collected, are more prone to this type of bias. Background noise at the time of the event as well as emotional context may affect memory and information processing. The QSF18 data can be used to provide insight into whether noise events are being emotionally encoded. The hypothesis being that if participants consistently change their reported annoyance level as time progresses between when they submitted the single event report and their daily summaries, then response latency has an impact on annoyance.

Approach

To assess consistency in sonic boom perception over time, a portion of QSF18 survey data were used that consist of instances where a participant responded to exactly one single event survey on a given day and completed the daily summary survey for the same day. The single event survey questions (and response options) of interest were: “Did you hear a sonic thump?” (yes or no) and “How much did the sonic thump bother, disturb, or annoy you?” (not at all, slightly, moderately, very, or extremely). The daily summary survey questions of interest were: “Did you hear any sonic thumps today?” (yes or no) and “Over the course of your day, how much did the sonic thumps bother, disturb, or annoy you?” (not at all, slightly, moderately, very, or extremely). The explicit assumption in this analysis approach is that the participant’s

single event response and daily summary response for these days describe their perception of the same sonic boom.

The following steps were used to narrow the data to instances where participants provided a single response to both surveys on a given day. First, the QSF18 data were subset into single event and daily summary response pairs with completed questionnaires from both surveys on the same test day. Responses were then removed for any of the following instances: a participant responded to a daily summary survey prior to a single event survey, a participant responded that they did not hear a sonic thump, and a participant responded to more than one single event survey in a day. The annoyance responses were recoded from the Likert-type scale, which ranges from “not at all annoyed” to “extremely annoyed”, into a numerical scale from 1 to 5 where 1 corresponds to “not at all annoyed” and 5 to “extremely annoyed.” This yielded a final sample of 186 pairs of one single event and daily summary questionnaires on the same test day.

Results

This section presents the assessment of perceptual consistency across reports using the subset described in the previous section. In addition, the effect of time between completion of single event and daily summary surveys is investigated to evaluate whether the amount of time impacts annoyance.

Figure 1 shows annoyance ratings for single event and daily summary pairs. A Sankey diagram in Figure 1(a) visualizes the few changes in annoyance rating between the single event and daily summary surveys. The width of the bands is proportional to the number of responses in a given response category. With relatively few departures from the initial rating, this diagram demonstrates that most participants reported consistently between the two surveys. Figure 1(b) depicts the same data but with numerical values. There are 167/186 = 89.8% responses that lie along the line y=x, again indicating that the responses are consistent.

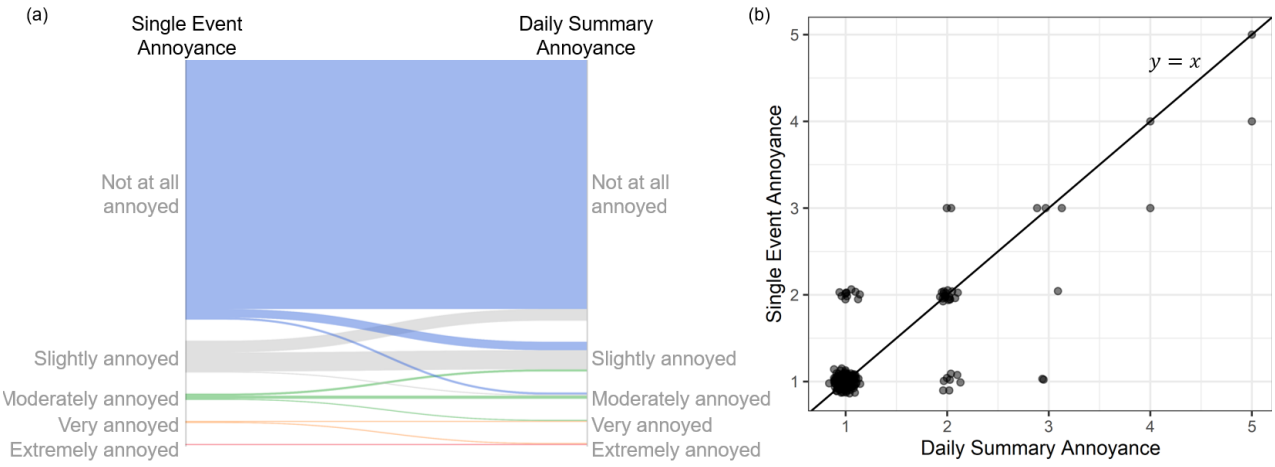


Figure 1. Trends in annoyance ratings between single event and corresponding daily summary surveys. (a) A Sankey diagram showing the flow of responses. (b) A scatterplot of the numeric responses.

Table 1 notes the number of instances of consistent and differing responses. Annoyance values from the daily summary were subtracted from single event responses. Therefore, 0

indicates no change, while a negative value indicates a decrease in annoyance and a positive value indicates an increase in annoyance. There are only 2 instances where responses differ by 2 and 17 that differ by 1. This shows that participants consistently report their annoyance to a sonic boom over the course of a day, and when they did not, the annoyance perception changed equally in both directions. Results suggest that annoyance did not shift in a particular direction as time passed.

Table 1. Summary of annoyance rating differences

Change in Rating from Single Event to Daily Summary	Count
Rating decreased by 1	9
Same rating	167
Rating increased by 1	8
Rating increased by 2	2

The effect of response latency is further tested as a potential indicator of a participant's proneness to report a higher annoyance value. Participants responded to single event surveys within one hour of a boom 75% of the time, though some responses were submitted many hours or even days later in a few cases. Figure 2 shows annoyance over time for those who submitted responses within 24 hours of the given boom. The y-axis shows annoyance from 1 to 5 and the x-axis shows survey completion latency, which is the amount of time between the single boom event and when a participant completes their single event report. Annoyance values of 5 were all submitted within 8 hours of the initial event. This analysis suggests that response latency does not affect annoyance perception of a sonic thump.

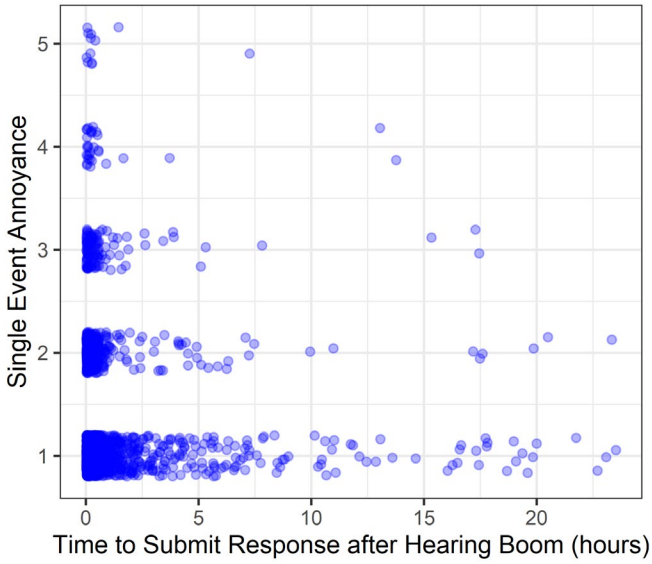


Figure 2. Single event survey completion latency within first day by annoyance responses (jittered to show more points)

The response latency is further explored via cumulative response curves in Figure 3. Cumulative responses are grouped by annoyance category and are shown as a function of time after a single event for the first 24 hours in Figure 3(a) and the first hour in Figure 3(b). A greater proportion of high annoyance (4 and 5) responses were submitted faster within the first hour than lower annoyance responses. This result supports the idea that if the thump is emotionally encoded as very or extremely annoying to participants, they may respond to the survey sooner.

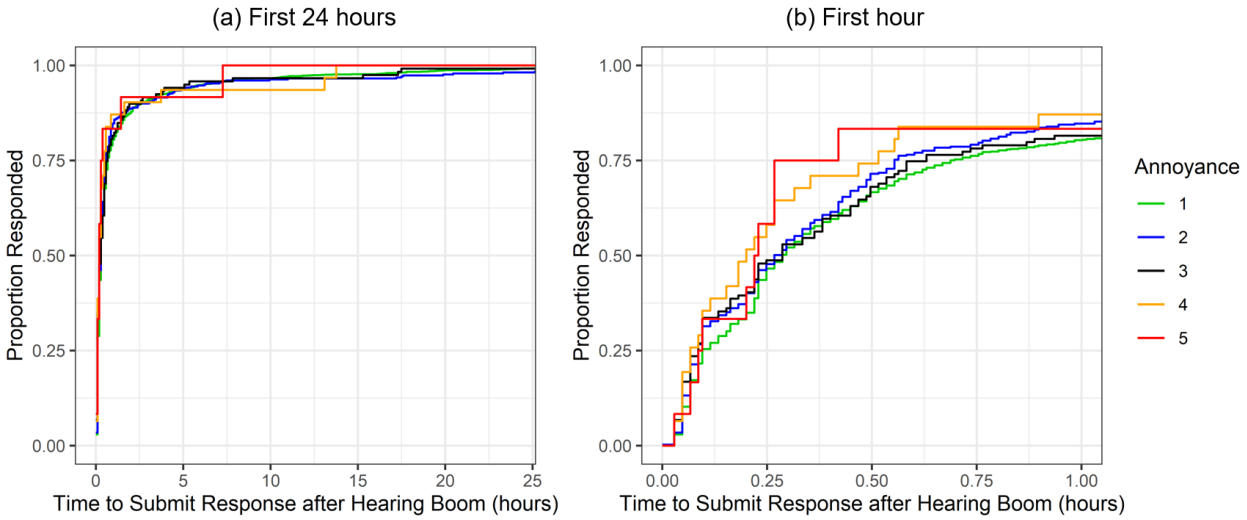


Figure 3. Cumulative responses received by annoyance rating within a) the first day and b) first hour of hearing the boom

The next step is to see if their response to the daily summary changes. The timing of booms and survey responses are further examined for their effect on recall in Figure 4. Figures 4(a) and 4(b) show the annoyance responses to single event and daily summary surveys versus the time between the boom and the response. Figures 4(c) and 4(d) show the difference in annoyance ratings for the same participant on the same day versus the time difference between the single or daily survey and the boom time. Figure 4(e) shows the difference between annoyance ratings by time difference to respond to the daily summary survey after the single event survey. The scatterplots show that there is little to no impact of timing on response differences over the course of the day. Both time from the boom and time since the single event survey do not appear to impact those who responded with different levels of annoyance between the two surveys. A more rigorous analysis may be warranted in the future to confirm this.

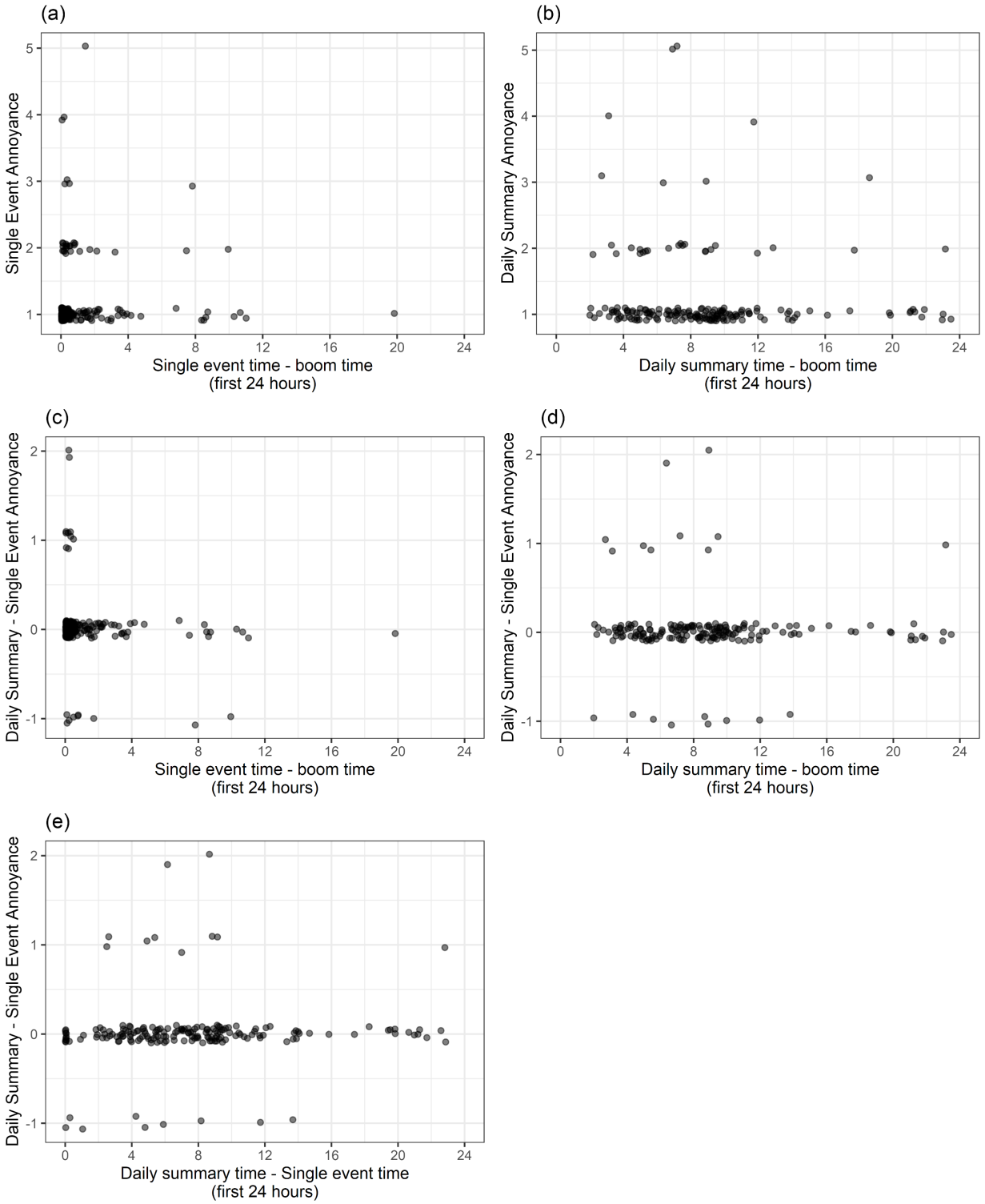


Figure 4. (a) Single event annoyance by time to respond after the boom. (b) Daily summary annoyance by time to respond after the boom. (c) Difference between single event and daily summary annoyance by time to respond to single event survey after the boom. (d) Difference between single event and daily summary annoyance by time to respond to daily summary survey after the boom. (e) Difference between single event and daily summary annoyance by time difference to respond to daily summary survey after single event survey.

Objective 2: Survey Satisficing with Early Termination

Satisficing is a common term in the field of survey methodology to explain a phenomenon that occurs when respondents "satisfice," i.e., take shortcuts to conserve effort or experience changes in motivation or ability. There are many ways a participant could satisfice including choosing the same response every time (i.e., "straightlining"), skipping, rushing, or exiting out of the survey without answering all questions (Barge and Gehlbach, 2012). This form of response saves the participant time and, in some cases, mental energy but is unideal for research purposes.

Objective 2 aims to answer the research question, "Will participants opt for a survey response that allows for early termination if the option to do so exists?" To answer this, the QSF18 dataset is examined for instances of early termination of the daily summary surveys by examining conflicting reports of hearing sonic booms throughout the day. For example, if in a given day a participant responded to at least one single event survey as having heard a sonic boom but indicated in the daily summary survey that they had heard no sonic booms, then the number of questions in the daily summary survey would be reduced, which could be evidence of satisficing.

Literature Review

Barge and Gehlbach (2011) identified prominent satisficing behaviors by administering two different surveys in higher education institutions in both Europe and the United States. The first survey contained 75 items, while the second contained 250. Researchers were looking for occurrences of early termination, non-differentiation, skipping items, and rushing. Early termination occurs when respondents stop taking the survey after completing a certain amount of it. Non-differentiation is a method of satisficing where respondents answer the same way for all questions. For example, if participants are given Likert response scales of "strongly agree" to "strongly disagree" they may simply choose "strongly agree" for each question without fully taking the time to read through the questions or consider their stance more carefully. Participants could also engage in survey satisficing by skipping items or rushing through the survey. They found that 61% of respondents demonstrated satisficing behaviors in the first survey, while 80% of respondents engaged in satisficing behaviors in the second survey. Most commonly, participants engaged in satisficing by skipping items in both surveys while rushing was the second most common behavior. Early termination and non-differentiation were third and fourth, respectively, for the first survey but vice versa for the second survey. The results of this work suggest that steps should be taken to mitigate these common satisficing behaviors in surveys through careful survey questionnaire design. This study also suggests what behaviors to search for in survey data as evidence of satisficing.

There are proposed methods in the literature to mitigate a respondent's willingness to engage in satisficing behaviors. Researchers posited that the arrangement of a rating scale on screen could affect response order effects[8] in web surveys (Höhne and Lenzer, 2015). This concept was measured with the use of eye tracking software as participants took a survey with

[8] Response order effects occur when the distribution of responses to a close-ended survey question is influenced by the order of response options provided.

Likert-scale responses arranged vertically and horizontally. Two groups were assigned to the two different arrangements of rating scales. The results demonstrated that response order effects were more likely to occur in vertical rather horizontal rating scales. Moreover, eye-tracking data also revealed that respondents do not read nor pay equal attention to all categories in a question. These results suggest that for a community sonic boom survey, a horizontal display of rating scales is preferred whenever possible; however, this might not be practical given the need for mobile phone survey administration (mobile phones are most commonly used in the "vertical" orientation). Furthermore, the ISO/TS 15666:2021 recommends vertical orientation for verbal-scale response choices in socio-acoustic surveys because it is not possible to evenly-space and easily view all of the response options when the scale is displayed horizontally on a phone (Clarke et al., 2021)

Another study evaluated context effects in web surveys and methods that can be taken to minimize them. Researchers employed a sample of 334 participants who were asked to participate in an online survey (Reips, 2002). Participants were sent a hyperlink that would intentionally buffer for some time before displaying the survey. This is known as a "high hurdle technique"[9] and was implemented to encourage those who are more likely to lack the patience to complete the survey to exit at the very beginning. This study provided insight into the effects of pop-up menus versus button scales and numerical labeling on survey satisficing behaviors. They found that these forms of administration did not have an impact on answering behavior of web participants. Lastly, the findings of this study indicate that the number of questions (i.e., one versus multiple) asked per page can cause a participant to produce different answers, which can be attributed to varying cognitive contexts evoked by the questions. This study suggests one question per page for surveys, as was done in the previous QSF18 community survey. It may be of interest in future community sonic boom surveys to employ a high hurdle technique to allow the sampled population to solely consist of those who are willing to fully complete the survey; however, it would need to consider the target sample size and the impact of this method on achieving the specified quota.

Another study in the same vein postulated that self-anchoring scales[10] could be used to lessen context effects in web surveys (Couper et al., 2004). Researchers in this study asked a sample of 266 students and non-students about their health with visuals to assess health. The experimental groups were presented with no picture, a negatively contrasting picture of a woman riding a bike, and a positively contrasting picture of a woman in a hospital bed. Negatively contrasting pictures were expected to enable participants to choose options on the scale indicative of poor health, while positively contrasting pictures were expected to influence them to rate their health with options indicative of good health. The results revealed that a larger proportion (27%) of respondents dropped off when using a self-anchoring scale compared to the 8% who quit while using the regular fixed anchors rating scale. While subjective preferences for either scale did not differ, early termination increased with self-anchoring scales. Verbal, fixed

[9] The high hurdle technique is a method that attempts to provoke participants with a low level of motivation to drop out early on in the study, with the remaining sample consisting of only highly motivated participants that are unlikely to drop off later (Reips, 2002).

[10] Self-anchoring scales are a type of response option that requires participants to assign their own definitions for the values along the scale.

anchors (i.e., “not at all annoyed” to “extremely annoyed”) have been utilized in previous community sonic boom surveys, and the work of Acker and Theuns (2009) supports its continued use in the future as the fixed anchors likely reduce early termination.

Another consideration in regard to eliminating context effects is the effect of participants’ age. An experiment conducted a secondary analysis on a sample of 777 respondents and asked two questions in different orders (Knäuper et al., 2007). Participants provided a yes or no response to the following questions: “Do you think it should be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children?” and “Do you think it should be possible for a pregnant woman to obtain a legal abortion if there is a strong chance of serious defect in the baby?” These questions were presented in different orders to evaluate context effects. The analysis demonstrated that the answers were less prone to change as age increased. They reported that 61% of respondents aged 18 to 54 supported abortion in the case of a woman who “does not want any more children” whereas only 49.6% did so when first presented with the “child defect” question. The discrepancy in percent support between question orders decreased as age increased until no order effect was observed for participants aged 65 and older. This study suggests that the 18 to 54 age group is more prone to having their responses influenced by their surrounding context. In terms of community sonic boom testing, including age as a covariate in dose-response modeling could test whether age is a significant factor and account for it if it is an issue. Speeding and straightlining through survey questionnaires may be an indicator of poor survey quality. In Zhang and Conrad (2013), a survey partially consisting of grid questions[11] was administered and survey responses were analyzed in order to correlate speeding and straightlining with various demographics. The definition of speeding used in the survey was 300 milliseconds per word, which was chosen to be slightly slower than the average college level reader. This definition was chosen with the simple rationale: selection of an answer faster than it is possible to read the question is a likely indicator of poor survey quality. Results showed that respondents sped on about 28% of the 54 questionnaire items. Their analysis confirmed that speeding was not an intermittent behavior and termed those who were prone to speeding as being “persistent speeders.” This behavior was noted to decrease with age, and a regression analysis confirmed that speeding and straightlining[12] in surveys were positively related.

However, there may be cases where speeding is not indicative of poor survey quality. Bassili and Fletcher (1991) found that participants who did not change their attitude when presented with a counterargument were more likely to respond to survey items faster. These rapid response times instead indicate simpler mental processes, which result in lower error rates. In the case of X-59 community surveys, while it is uncertain what the exact implications are of participants speeding through surveys, it is important to design the survey to reduce the potential for satisficing where possible in order to minimize the potential for survey error.

---

[11] From Zhang and Conrad, a grid question is “a type of survey question in which multiple items with the same response scales are displayed in a table-like format in which the items – usually statements – are the rows and the response categories appear as the columns.”

[12] Straightlining, also called non-differentiation, is a term used to explain a phenomenon in which a survey respondent chooses the same answer across an array of questions given the same response scale in order to preserve time and mental effort.

Approach

Survey data collected at QSF18 can be assessed for evidence of satisficing. While QSF18 questionnaires follow many of the measures suggested by the literature to mitigate common satisficing behaviors, satisficing is possible by using a shortcut to rush through the survey. Both the single event and daily summary surveys contained a question that asked whether participants heard a sonic boom. If the participant selected that they did not hear the boom, the survey would skip several questions. This leaves the possibility that if participants had mind to satisfice, then they would misreport as having not heard the boom. The objective is to evaluate whether participants had consistently reported having not heard a thump in the daily summary questionnaire even though they responded as having heard a single event thump during the same day.

The data for this analysis consist of instances where a participant responded that they did hear the boom on at least one single event survey on a given day and completed the daily summary survey for the same day. Cases where a participant reports that they had not heard any thumps in the daily summary survey are referred to as “inconsistent reports”.

Results

Survey satisficing is examined by comparing the response to the thump audibility question on single event and daily summary surveys. The daily summary survey completion time does not appear to change substantially with additional questions for those who reported that they did hear a boom that day. Ninety percent of participants submitted their daily summary survey within 3 minutes of starting it, regardless of whether they reported hearing any booms. Figure 5 displays the distribution of daily summary completion times and is subdivided by those who completed the survey within 10 minutes of starting and those that took longer than 10 minutes to complete it, as well as by whether or not a boom was reported as heard. It does not appear that time to complete the survey increased significantly if participants reported hearing the booms and answered the additional questions.

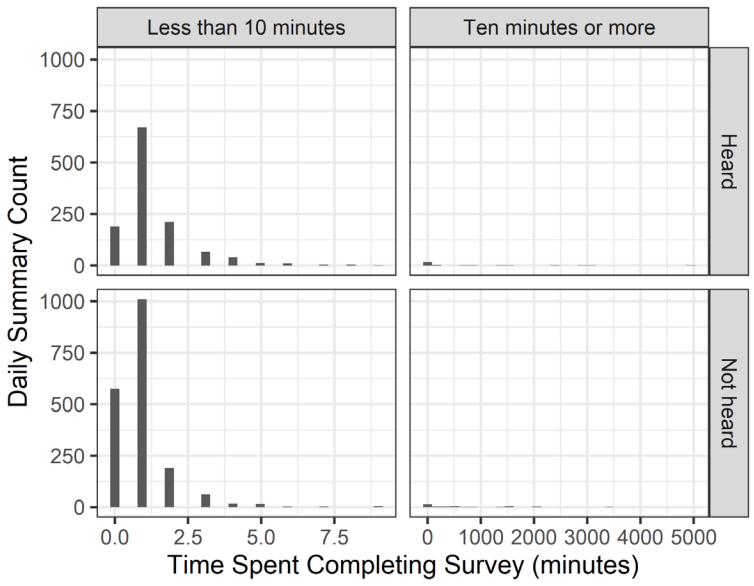


Figure 5. Daily summary survey duration distribution

Pairing daily summary and single events responses allow for inconsistency checks at an individual level. Across the nine test days there were a total of 66 inconsistent reports out of 767 complete pairs of single event and daily summary surveys, from just 32 participants. Figure 6 shows the number of inconsistent reports per participant. Of those who did have inconsistent reports, the majority (25/32 = 78.1%) only had one inconsistent report and only a few participants persistently misreported hearing booms on the two surveys. There are inconsistent reports for five participants (15.6%) on two test days, one participant (3.1%) on three test days, and one participant (3.1%) on four test days. With inconsistent reports from only 32 of 371 participants, a relatively small percentage (8.6%) of participants potentially engaged in satisficing behaviors. This supports the overall notion that participants are not satisficing by inconsistently responding to the thump audibility question in order to complete the survey faster

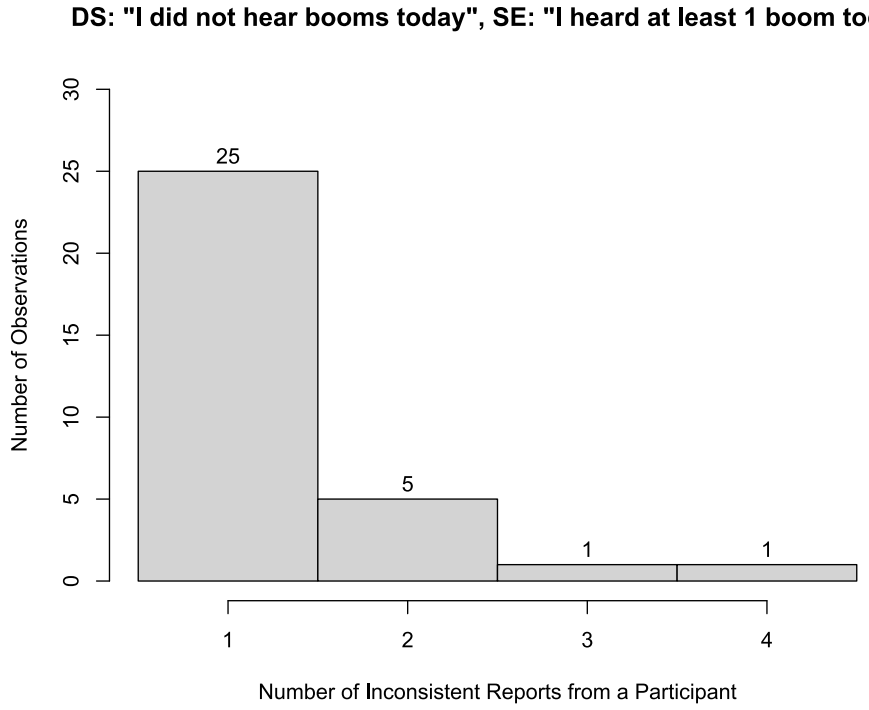


Figure 6. Number of Inconsistent Reports from a Participant Across Test Days according to Daily Summary (DS) and Single Event (SE) surveys.

Conclusion

There is a potential in community sonic boom tests for survey measurement error due to recall errors. This occurs when a reported value differs from a participant's true experience. The two topics examined in this paper are 1) how consistently participants recall their annoyance to sonic thump events and 2) do participants opt for early survey termination if the option is available. Data from QSF18 were examined for inconsistencies between single event and daily summary survey reports. Identification of inconsistent reports provides insights into the possible source of error.

A subset of QSF18 data paired single event and daily summary surveys to investigate perceptual consistency. The daily summary annoyance ratings matched the single event annoyance ratings exactly for the majority of instances. For the daily summary responses that did not match their single event counterparts, there was an equal change to increased and decreased annoyance ratings. This suggests that sonic thump annoyance recollection is consistent for most people over a few hours when comparing responses to a single heard boom event on a given day. There is no evidence to suggest that annoyance changes in any particular direction with time. These findings do not offer any insight into the optimal response interval between when a participant heard the sonic boom and when they submitted a response since the cases where participants did change their answer were in highly varied timeframes. Additionally, there are no trends or possibly not enough data to claim a need for an optimal response time.

Another source of measurement error is survey satisficing, or early termination. As with recall, concerns with satisficing are that participants simply chose an answer for a reason other than their own perception of the construct. For example, if a participant rushes to complete a survey because they want to simply get it done as fast as possible, then they may not necessarily select the response that most accurately describes their experience or attitude. In the QSF18 survey, participants could shorten their daily summary by reporting that they did not hear a boom that day, even if they had responded as having heard a boom in a single event survey that day. Instances of participants inconsistently reported boom audibility between the two surveys are relatively few. Moreover, the misreports may not have necessarily been due to participants satisficing but just accidentally selecting the wrong option on a survey. This is supported by the fact that nearly all participants who misreported only did so once or twice.

Research-backed methods for survey construction and administration are vital in ensuring data quality at the end of a study. In attempts to identify contributors to error, this work has demonstrated support for the continued use of numerous practices from the previous community tests. To an extent, the results of this study provide evidence to rule out recall bias and satisficing as sources of error.

Limitations and Future Work

The initial literature review provided many ways to approach survey measurement error that could be applied to the QSF18 data; however, not all were considered in this initial analysis. Some items not comprised in this paper include the influence of varying amounts of time between sonic boom exposure and the participant's single event survey response, time between sonic boom exposure and participant's daily summary survey response, and the time between a participant's single event survey response and daily survey response. The assumption that a participant only heard one boom for the perceptual consistency analysis may not hold as it is possible a respondent heard more than one sonic boom during the day, but only completed a single event survey for one of the booms. Including the estimated cumulative dose could help confirm whether a participant was exposed to only one boom. The present analysis also ignored any impact of the loudness of the single event sonic boom.

Future work in regard to survey satisficing could include evaluating whether participants engage in any satisficing behaviors beyond early termination or rushing. Examples of possible

behaviors include straightlining responses or skipping items. If there were participants who responded that they had heard the thump but provided no response for annoyance, then this could indicate that participants simply lost interest in completing the survey. It would require in-depth reasoning to identify this behavior as survey error as opposed to a crystallized attitude.

The survey completion time can be further investigated for potential speeding and possibly determine an optimal response time. Since annoyance remained constant for most participants, it is possible that some participants may have been speeders who answered questions the same way on all surveys throughout the test. Analyzing the survey time duration for participants who responded with the same annoyance in each instance could address this issue. The objective would be to identify an average time for survey completion and consider whether consistent reporting is a byproduct of speeding. If this analysis reveals that the earliest responses instead have the most variability in their answers, then perhaps there is an opportunity to identify an optimal survey response time. If speeders are identified, it may be worth removing speeders to ascertain overall perception of accuracy without context effects.

References

Anikin, A. (2020). The link between auditory salience and emotion intensity. *Cognition and Emotion*, 34(6), 1246–1259. doi: 10.1080/02699931.2020.1736992

Barge, S., and Gehlbach, H. (2011). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, *53*(2), 182–200. doi: 10.1007/s11162-011-9251-2

Bassili, J. N., and Fletcher, J. F. (1991). Response-time measurement in survey research a method for CATI and a new look at nonattitudes. *Public Opinion Quarterly*, 55(3), 331-346. doi: 10.1086/269265

Brierley, B., Medford, N., Shaw, P., and David, A.S. (2007). Emotional memory for words: Separating content and context. *Cognition and Emotion,* 21(3), 495-521. doi: 10.1016/j.pscychresns.2004.10.004

Couper, Mick P., Roger Tourangeau, and K. Kenyon. 2004. "Picture This! Exploring Visual Effects in Web Surveys." *Public Opinion Quarterly*, 68(2): 255-266. doi: 10.1093/poq/nfh013

Clarke, C., Gjestland, T., Lavia, L., and (2021). Assessing community noise annoyance: A review of two decades of international technical specification ISO/TS 15666:2003. *Journal of the Acoustical Society of America*, 150, 3362–3373. doi: 10.1121/10.0006967

Clarke, P.M., Fiebig, D.G., and Gerdtham, U. (2008). Optimal recall length in survey design. *Journal of Health Economics*, 27(5), 1275-84. doi: 10.1016/j.jhealeco.2008.05.012

Davenport, W.G. (1972). Vigilance and arousal: Effects of different types of background stimulation. *The Journal of Psychology*, *82*(2), 339–346. doi: 10.1080/00223980.1972.9923824

Haering, E. A., Jr., Smolka, J. W., Murray, J. E., and Plotkin, K. J. (2006). “Flight demonstration of low overpressure n-wave sonic booms and evanescent waves,” AIP Conf. Proc. 838, 647–650. 10.1063/1.2210436

Höhne, J., and Lenzner, T. (2015). Investigating response order effects in web surveys using eye tracking. *Psihologija*, 48, 361-377. doi: 10.2298/PSI1504361H

ISO (2021). ISO/TS15666:2021, Acoustics—Assessment of noise annoyance by means of social and socio-acoustic surveys. International Organization for Standardization, Geneva, Switzerland.

Knauper, B., Schwarz, N., Park, D., and Fritsch, A. (2007). The perils of interpreting age differences in attitude. *Journal of Official Statistics*, *23*, 515-528.

Krosnick, J.A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-136. 10.1002/acp.2350050305

Lee, J., Rathsam J., and Wilson, A. (2019). *Statistical modeling of quiet sonic boom community response survey data*. NASA/TM–2019–220427. ntrs.nasa.gov/citations/20190033466

MacLeod, C.M., and Nelson, T.O. (1984). Response latency and response accuracy as measures of memory. *Acta Psychologica, 57*(3), 215–235. doi: 10.1016/0001-6918(84)90032-5

Reips, U. (2002). "Context effects in web surveys," *Online Social Sciences*, Hogrefe and Huber, 69-80. doi: 10.5167/uzh-19732

Van Acker, F., and Theuns, P. (2009). On possibilities and limitations of using self-anchoring scales in web surveys. *Quality and Quantity, 44*(6), 1129–1137. doi: 10.1007/s11135-009-9265-4

Zhang, C., and Conrad, F. (2013). Speeding in web surveys: the tendency to answer very fast and its association with straightlining. *Survey Research Methods, 8*, 127-135. doi: 10.18148/srm/2014.v8i2.5453