Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

A paired comparison experiment to examine startle evoked by low level sonic booms and other transients.

Andrew Marshall (1) and Patricia Davies (1)

(1) Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, West Lafayette,IN ,USA

PACS: 43.28.MW, 43.50.PN, 43.66.LJ

ABSTRACT

One important aspect of the operation of next generation supersonic aircraft is the potential impact that low amplitude sonic booms will have on people. Due to the quick rise of these sounds, startle responses are possible. In two previous semantic differential experiments, judgments of startle were found to be highly correlated with judgments of annoyance. In addition, judgments of loudness could not fully explain startle or annoyance ratings. The linear models predicting startle or annoyance that performed best were based on the maximum loudness and maximum loudness derivative, as calculated by using Glasberg and Moore's time-varying loudness algorithm. Research has been focused on improving this model of startle and to examine how physiological responses relate to subjects' ratings of startle. As part of an experiment designed to look at repeatability of subjects' physiological responses and to examine more carefully the influence of loudness derivative on annoyance, a paired comparison test was designed. The maximum loudness and loudness derivative of the five low level sonic boom stimuli were controlled to cover a range in which the threshold where physiological responses associated with startle is found. Subjects completed two sessions, each 24 hours apart and in each session the paired comparison test was repeated three times. In each of the six paired comparison tests, subjects heard the 20 pairs of sounds and selected which sound was more startling. The repeatability of subject judgments across all six paired comparison tests is discussed as is the impact of loudness derivative on the judgments of the sounds.

INTRODUCTION

Due to fast rise times and relatively large peak pressure, one important aspect of the impact of sonic booms on humans is that of startle. Previous studies have shown that people's evaluations of the startle evoked by low level sonic booms could not be fully explained by loudness alone (Marshall and Davies, 2007 and 2008). In addition, judgments of annoyance were highly correlated to judgments of startle in both of those studies. Metrics to predict subject-rated startle and annoyance were also examined in these studies. The models which performed best were based on a linear combination of maximum loudness and other statistics of the output of Glasberg and Moore's time-varying loudness (2002). However, the analysis of linear models of two or more metrics in this previous work was exploratory; many of the additional statistics of time-varying loudness used were correlated to one another and to maximum loudness. In addition it was unclear both how subjective startle is related to physiological responses associated with startle. It was also unknown how subjective startle judgements vary both over the course of an experiment and from day-to-day.

A paired comparison experiment was designed as part of a larger test created to examine the repeatability of subjects' physiological response to sonic booms. Subjects completed the experiment twice, each on on different days to examine how responses varied from day-to-day.

EXPERIMENTAL METHODOLOGY

The test involved subjects resting their chin on a headrest looking at pictures, listening to sounds and choosing one of a pair of sounds based on which they found more startling. Subjects were instrumented with three sets of measures: electromyography, skin conductance and finger pulse. The details of the experimental procedure are described below.

Procedure

IRB approval was obtained for this experimental protocol (Protocol #0904007961). First the experiment was briefly explained to the subjects. Then, informed consent was obtained and the subject's hearing tested. The subject was then placed into the sound booth and the finger cuffs (two for measurement of skin conductance one for measurement of pulserate) and EMG electrodes were attached. At this time, the headrest was adjusted by the subject for comfort. Next, the subject completed a calibration procedure, which involved sitting relaxed with and without the chin-rest, doing head rotations, swallowing, and coughing. These particular movements in this procedure produced movement artifacts in previous experiments (Marshall, Davies and Rietdyk, 2009). To ensure consistency in carrying out this procedure across sessions, the instructions for the calibration procedure were pre-recorded.

Once the calibration was complete, the subjects completed an experiment where they viewed pictures of plants and landscapes while being exposed to the experimental stimuli. The subjects were exposed to each of the five stimuli twice over the course of the 25 minute test, with roughly 2 minutes between exposures. The goal of this part of the experiment was to examine the repeatability of physiological responses. Preliminary results of this work are presented in Marshall and Davies (2010).

After completing the picture-portion of the experiment, the subject completed the paired comparison experiment using the same stimuli. For each trial, the subject heard a pair of sounds with 1 second delay between each sound and was asked to determine which sound was "more startling". In this part, the order of signal pair exposures was randomized with a different order for each session. Each subject completed the paired comparison test 3 times per session. In each test they were presented with 20 pairs of sounds, all pair combinations consisting of different signals and subjects heard each pair in both orderings: (I,J) and (J,I), at some time in the test. Then the subject's hearing was tested again, and the subject was compensated and thanked for volunteering. The next day, the subject repeated the experiment again, following the same procedure outlined above.

Subjects

Thirteen subjects, 8 male and 5 female were recruited for this study. Subject ages ranged from 24-54 years and all were students recruited from the area surrounding Purdue University. All subjects had less than 20 dB of hearing loss between the frequencies of 125-8000 Hz. Each subject completed one session of this experiment per day over the course of two days. Subjects were paid $10 for participating in each session.

Apparatus

The sounds were presented to each subject in an IAC double-walled sound booth using Etymotics ER-2 earphones. Signals were presented binaurally and pre-filtered to account for the operating mode of the ER-2 earphones. As mentioned earlier, to minimize subject head movement, which causes movement artifacts in the physiological measures, the subjects rested their heads on a chin rest, attached at a comfortable distance from a computer screen.

During the picture-viewing portion of the experiment, skin conductance, pulse rate and electrical activity of the neck muscles (EMG) were measured. These measures were not analysed for the paired comparison experiment due to the short interstimulus interval which caused physiological response overlap.

Signals

The experimental stimuli were five synthetic sonic booms generated by using the procedure outlined in Marshall and Davies (2007b). Briefly, signals were generated by frequency domain sampling, a technique commonly used in finite impulse response (FIR) filter design, where the spectrum originated from the Fourier transform of a candidate continuous signal. The frequency domain technique was used to avoid aliasing induced by sampling the candidate signal in the time-domain.

Because the main purpose of this experiment was to examine the repeatability of startle responses, it was necessary to select signals that were likely to cause startle. Thackray, Touchstone and Bailey (1974) used booms of 16, 30 and 50 Pa peak overpressure and observed statistically similar startle responses evoked by the booms of 30 and 50 Pa peak overpressure. However, due to limitations in playback, Thackray *et al.* (1974) did not vary the rise time of the stimuli. Because the time-histories of low-level shaped sonic boom vary substantially from the classical N-waves used in Thackray *et al.*'s study, it was necessary to normalize the stimuli of this experiment to produce similar loudness levels to those for sounds in that experiment. To do this the maximum loudness of the 16 and 30 Pa booms that Thackray *et al.* (1974) were calculated by using Glasberg and Moore's time-varying loudness (Glasberg and Moore, 2002). In addition, the maximum loudness derivative was also calculated for the period from the start of the signal to when the first peak in the loudness time-history occurred. The derivative was calculated by using a 50 pt finite impulse responses differentiator filter. The five stimuli of this experiment were designed so that they had approximately the same maximum loudness as these two sounds and others had maximum loudness derivatives that bracketed the loudness derivative of those sounds (Figure 1).

Due to the limitation in the playback system to reproduce low frequencies, all signals were high passed filtered. Signals 1,4 and 5 were high pass filter with a cut-off of at least 25 Hz and signals 2 and 3 were high-pass filtered with a cut-off of 35 Hz. The higher cut-off frequency for signals 2 and 3 was necessary to play these signals at the correct maximum loudness. This filtering was accounted for in the calculation of metrics.

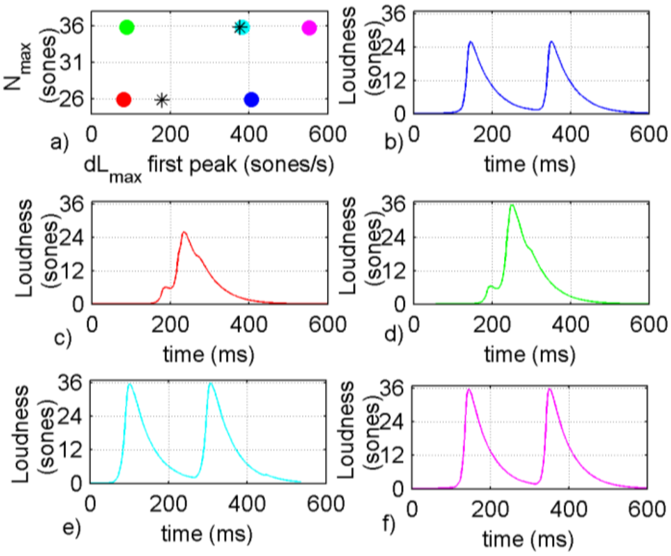

Figure 1: a) Maximum loudness (sones) plotted versus maximum loudness derivative (sones/sec) for test stimuli (circles) and for Thackray *et al.*'s stimuli (star). Loudness time-histories for b) signal 1 (blue), c) signal 2 (red), d) signal 3 (green), e) signal 4 (cyan) and f) signal 5 (magenta).

RESULTS

Ratings

The data from paired comparison tests were converted into probabilities and transformed into startle scores based on the Bradley Terry Luce [BTL] model (Guilford, 1954). To obtain estimates of the variability of each score, a series of probability matrices were generated, each was randomly permuted in the range of +/- 2 standard deviations of the original probability matrix created from subjects' responses. Estimates of the standard deviation of the scores were obtained by taking the standard deviation of the BTL scores obtained from the series of the permuted probability matrices. To examine the effect of individual subjects on the startle scores, BTL scores were calculated for the case when each individual subject was removed from analysis. To aid in interpretation, all scores were normalized to have a score of zero for signal 1. The results of this analysis are plotted in Figure 2.

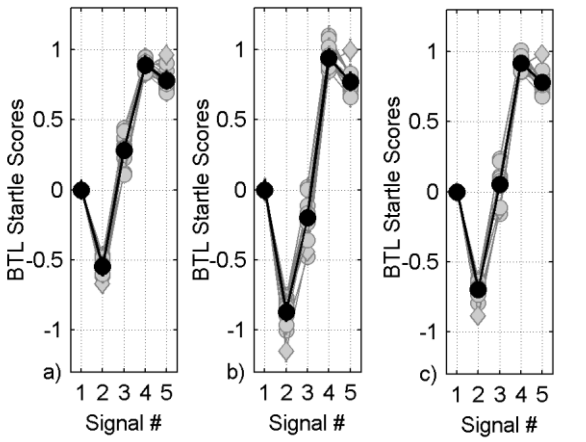

Figure 2: BTL startle scores for a) first day, b) second day and c) both days. Black is all subjects. Gray corresponds to scores calculated when each subject's responses are eliminated from the group. The gray scores with diamonds is the result when subject 9 was removed.

In general, the removal of an individual subject's responses from the group results in small changes to the BTL scores. The largest changes in scores occurrs when subject 9 is removed (diamond), but this difference was relatively small, hence all subjects are retained for analysis. The BTL scores for signals 2, 3 and 5 have the largest subject-to-subject and day-to-day variations in BTL scores. These signals were those with either the smallest or the largest maximum loudness derivative (of the first peak of the signal) [dLmax]. The larger span of BTL scores on the second day is primarily due to relatively lower BTL scores (compared to signal 1) of signals 2 and 3; the scores of the other signals are relatively unchanged. It is possible that these signals produced very low levels of startle; repeated exposures results in reduced sensitivity or greater habituation and thus does not appear to occur with other, more startling sounds.

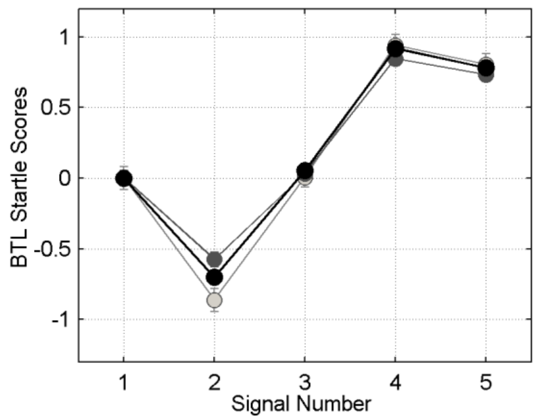

Figure 3: BTL startle scores for first test (light gray), second and third tests (dark gray) and all tests (black) for both days.

The BTL scores from 1st test on each day were compared to those from the 2nd and 3rd tests on the same day (Figure 3). If repeated exposure to signals 2 and 3 resulted in relatively smaller BTL scores due to habituation, smaller scores are expected from the 2nd and 3rd test. However, this did not occur. There is very little variation between the first and the other tests for the scores of signal 3 and the signal 2 scores increase in the latter tests. There does appear to be a slight decrease in the range of the of BTL scores particularly for signals 2 and 4, but this effect is very small. This lack of variation is not unexpected considering that the subjects were exposed to the signals just before during the picture-viewing poriton of the experiment that occurred before the paired comparison test.

Metrics: Linear Models

To examine if the responses are related to noise metrics (maximum loudness and dLmax), the BTL startle scores are plotted against each of these metrics in Figure 4. Both metrics individually predict the startle scores similarly well (R^2=0.61 versus R^2=0.58). Maximum loudness is a slightly better predictor, but this result is extremely sensitive to variations in the startle scores. BTL scores calculated by removing subject 9's responses, for example, results in R^2 values of 0.52 and 0.71 for maximum loudness and dLmax, respectively. This result is primarily due to the relative position of signals 3 and 5. Of importance is that signals 1 and 3 appear to have nearly the same BTL startle scores, despite having very different maximum loudnesses. From this result, it appears that an additional 300 sones/s of dLmax produces an effect equivalent to that produced by a change of 10 sones in maximum loudness. When compared to Marshall and Davies (2007), dLmax is more strongly correlated to startle scores in this test. However, this may be expected because of the larger variation of dLmax in the experimental stimuli.

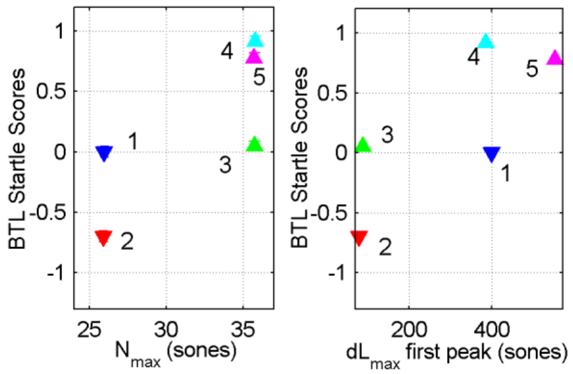

Figure 4: BTL startle scores versus maximum loudness (R^2=0.61)(left plot) and maximum loudness derivative of the first peak (R^2=0.58)(right plot). Signals are color coordinated to match colors in Figure 1.

Although there are only five signals in this experiment, it is possible to investigate how well some of the linear models examined in previous research (e.g. Marshall and Davies, 2007) predict the BTL startle scores. Aside models that include dLmax as a factor, models incorporating maximum loudness with either the reciprocal of loudness rise time (1/MRT) or mean Zwicker Sharpness (Smean) were examined. Loudness rise time was defined as the time between where the loudness time-history breached the noise floor to the point where the first loudness peak was reached. For each linear model, the parameters were estimated from the startle scores in the study reported in Marshall and Davies (2007), which ranged from -16 (Calming) to +16 (Startling). Then these linear models were used to predict the BTL startle scores and coeficents of determination were calculated (R^2). The models that included Smean or 1/MRT predicted BTL values worked almost as well as the Maximum loudness or dLmax alone (R^2=0.55 and 0.62 for the models that included Smean and 1/MRT, respectively). The linear model incorporating maximum loudness and dLmax performed better (R^2=0.88) and BTL scores are plotted versus predictions in Figure 5.

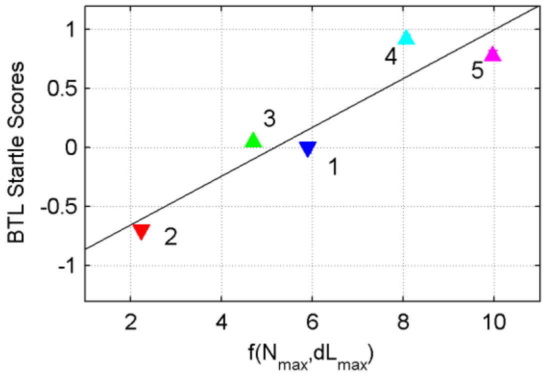

Figure 5: BTL startle scores versus linear model incorporatingg maximum loudness and dLmax (R^2=0.88). Signals are color coordinated to match colors in Figure 1.

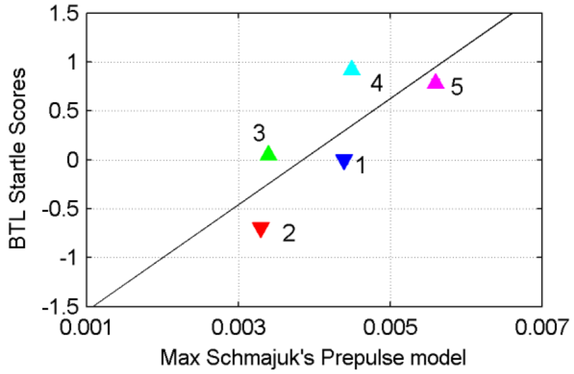

Figure 6: BTL startle scores versus maximum of the output of Schmajuk's prepulse inhibition model (R^2=0.60). Signals are color coordinated to match colors in Figure 1.

The poor performance of the other linear model predictions, is probably due to the lack of variation in the loudness rise time and sharpness in this experiment compared with the variation in stimuli used in Marshall and Davies (2007).

Prepulse model

One important aspect of startle is prepulse inhibition and facilitation (Schmajuk and Larrauri, 2006). Briefly, this phenomenon is when the presence of a sound event (prepulse) precedes a startling stimulus (pulse). Depending on the magnitude of the prepulse and the interval between the pulse and the prepulse the startle evoked by the main event can be attenuation or enhanced. Because shaped sonic booms can have large asymmetries in the pressure signature, prepulse inhibition may be a strategy to minimize startle evoked by low booms.

To examine the effect of prepulse inhibition on the stimuli, the model of Schmajuk and Larraui (2005) was used. The input to this model is the instantaneous level (in dB) of white noise as measured by a sound level meter and sampled at 1000 Hz. To create stimuli that could be used with this model, a series of windowed white noise signals were generated to have the same loudness time histories as the paired comparison test stimuli. This was accomplished by iteratively adjusting the level of the white noise in 1millisecond intervals. Then, to dertmine the input to the prepulse model, the fast-average sound pressure level was calculed from the white noise surrogate signals.

By using this procedure, the maximum error in the instantaneous loudness time histories was 1 sone (< 2% of the instantaeous loudness of the booms). The output of the prespulse model was obtained and maximum determined. The BTL scores versus the maximum output of the prepulse model are plotted in Figure 6.

Schmajuk's model performs about as well as the other linear models and dLmax alone. Likely dLmax, the prepulse model underpredicts the score of signal 4 and overpredicts the score of signal 1. Interestingly, the maximum of Schmajuk's model is strongly correlated to dLmax (R=0.97). This result is not surprising; the structure of Schmajuk's model consists of segments where the rise of the input is compared to a "running average" of the input. Thus, signals with quicker rise in level would result in large outputs.

CONCLUDING COMMENTS

As part of an experiment designed to investigate the repeatability of physiological responses to low amplitude sonic booms, a paired comparison test was conducted. Subjects completed a total of 6 tests, 3 on each day. It was found that BTL scores varied the most for the signals with the smallest maximum loudness derivative [dLmax] of the first part of the signal. Previously developed linear models of startle based on maximum loudness and either mean Zwicker sharpness (Smean), dLmax and loudness rise time (MRT) were also examined. It was found that the linear model consisting of dLmax and maximum loudness was the best predictor of BTL startle scores. However, the performance of these other models might have been better if the stimulus set had contained more variation of those sounds characteristics. Schmajuk's prepulse inhibition model was investigated and its maximum output was found to be strongly correlated with dLmax. However, evaluations of all of the models' performance were limited by the small number of stimuli in this experiment. Of note here is that the rate of changes of loudness is important to consider along with peak loudness when considering how people may be impacted by low amplitude sonic booms.

ACKNOWLEDGEMENTS

The authors want to thank Debra Arihood for giving us a database of photos to use for this research. The authors also wish to thank Professor Shirley Rietdyk of Purdue University for advice regarding the electromyography portions of this experiment. The authors also wish to thank the FAA/NASA/TC PARTNER Center of Excellence for their financial support. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the FAA, NASA or Transport Canada.

REFERENCES

1. M.M. Bradley and P.J. Lang. Affective reactions to acoustic stimuli. *Psychophysiology*, 37:204-215, 2000
2. R. Glasberg and B. C. J. Moore. A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, 50(5):331–42, May 2002.
3. Guilford, *Psychometric Methods*, McGraw-Hill, 1954.
4. Marshall and P. Davies, A semantic differential study of low amplitude supersonic aircraft noise and other transient sounds, *Proceedings of the International Congress on Acoustics*, Madrid, September, 2007. Paper No. NL8_ICA_6_am
5. A. Marshall and P. Davies, On the use of statistics of a time-varying loudness model to quantify loudness of

transient environmental noise, *Proceedings of Noisecon 2007*, Reno, NV, US, October, 2007.

6. A. Marshall and P.Davies, A comparison of perceptual attributes and ratings of environmental transient sounds in two different playback environments. *Proceedings of Internoise 2008,* Shanghai, China, October 2008. Paper No. in08_0869
7. A. Marshall, P. Davies, and S. Rietdyk, An investigation of differences in startle physiological measures between sonic booms and other transients, *Proceedings of Internoise 2009*, Ottawa, Canada, August 2009.
8. A. Marshall and P. Davies, An examination of repeatability in physiological responses to simulated low amplitude sonic booms of various peak levels and rise times. Joint ASA/NoiseCon2010, Baltimore, MD, 2010.
9. R.I. Thackray, R.M Touchstone and J.P Bailey, Behavioral Autonomic and subjective reactions to low- and moderate-level simulated sonic booms: A report of two experiments and a general evaluation of sonic boom startle effect, FAA technical report: FAA-AM-74-9, 1974.
10. N. Schmajuk and J. Larrauri. *Neurotransmitter interactions and congnitive function*, chapter: Prespulse inhibition mechanism and congitive processes: a review and model, pages 245–278. 2006.
11. Schmajuk and Larrauri, A neural network model of prepulse inhibition, *Behavior neuroscience*,119(6):1546-1562, 2005.