Loudness and short-term annoyance of sonic boom signatures at low levels
Stephan Töpken, and Steven van de Par

Citation: The Journal of the Acoustical Society of America 149, 2004 (2021); doi: 10.1121/10.0003779
View online: https://doi.org/10.1121/10.0003779
View Table of Contents: https://asa.scitation.org/toc/jas/149/3
Published by the Acoustical Society of America

ARTICLES YOU MAY BE INTERESTED IN
Allan Pierce and adiabatic normal modes
The Journal of the Acoustical Society of America 149, R5 (2021); https://doi.org/10.1121/10.0003595

Closed-aperture unbounded acoustics experimentation using multidimensional deconvolution
The Journal of the Acoustical Society of America 149, 1813 (2021); https://doi.org/10.1121/10.0003706

Annoyance, perception, and physiological effects of wind turbine infrasound
The Journal of the Acoustical Society of America 149, 2238 (2021); https://doi.org/10.1121/10.0003509

Acoustics modelling of open-cell foam materials from microstructure and constitutive properties
The Journal of the Acoustical Society of America 149, 2016 (2021); https://doi.org/10.1121/10.0003824

Vertical and slanted sound propagation in the near-ground atmosphere: Amplitude and phase fluctuations
The Journal of the Acoustical Society of America 149, 2055 (2021); https://doi.org/10.1121/10.0003820

A method for realistic, conversational signal-to-noise ratio estimation
The Journal of the Acoustical Society of America 149, 1559 (2021); https://doi.org/10.1121/10.0003626
Loudness and short-term annoyance of sonic boom signatures at low levels

Stephan Topken and Steven van de Par

Acoustics Group, Department of Medical Physics and Acoustics, Excellence Cluster Hearing4all, Carl von Ossietzky University, Carl-von-Ossietzky-Strasse 9-11, 26129 Oldenburg, Germany

ABSTRACT:
Supersonic aircraft produce a sonic boom when flying faster than the speed of sound. To rule out detrimental effects for inhabitants of overflown areas, civil supersonic flights (like the Concorde) were only allowed to fly at supersonic speed when over water. Due to the progress in aircraft design, the supersonic boom may be reduced considerably in the future. In this study, listening tests were carried out with a variety of low boom simulations and conventional sonic boom signatures in a similar level range. Participants rated the loudness and the short-term annoyance of 24 sonic boom signals, which differed in terms of the signature shape and maximum pressure but were confined to a range of A-weighted sound exposure levels around 60 dB(A). The results showed main effects of signature and relative level variation as well as an interaction of the two. Correlation coefficients between the ratings and sound exposure levels were highest for A-weighted sound exposure levels compared to other frequency weightings. Contrary to our expectations, the provision of background information about the nature of the presented sound sources had no statistically significant influence on the ratings.

© 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1121/10.0003779

(Received 17 December 2020; accepted 25 February 2021; published online 24 March 2021)

[Editor: Sanford Fidell]

I. INTRODUCTION

The sonic boom level of future supersonic aircraft may be reduced considerably due to the progress in aircraft design. Such low sonic boom sounds will be considerably quieter and sound very different compared to conventional sonic booms. The evoked sensation and subjective response of human listeners to future low sonic boom signatures are under investigation. However, a definitive acoustic measure reflecting the acceptability of low sonic booms is not established yet, and there is a question of how to define it such that the effects on humans are well reflected.

Several studies investigated the loudness, the loudness in combination with acceptability, the annoyance, and the realism of sonic boom sounds. In search of an acoustic measure reflecting human perception, some studies found close correlations between sonic boom ratings and values of the A-weighted sound exposure level (ASEL). Leatherwood and Sullivan investigated the effect of boom shapes for simple simulations constructed from connected straight lines. They found correlation coefficients around $r = 0.96$ between the ASEL and loudness ratings for overpressure values between 50 and 125 Pa (1.0 psf and 2.6 psf) and ASEL values between 62 and 90 dB(A), depending on the set of sounds. The same authors also reported very high correlation coefficients between ASEL values and loudness/annoyance values obtained from magnitude estimation experiments.

More recent studies found similar values for coefficients of determination between annoyance ratings and ASEL or perceived level (PL) for simulated low intensity sonic booms and for N-waves with different rise times and filtered impulse signals. Similar results were found for sonic boom sounds with and without rattle when calculated for the actual sound heard by participants (e.g., indoors). Vos investigated the annoyance produced by small, medium-large, and large firearms. He achieved a more precise and unified characterization of the annoyance from impulsive events by using a rating sound level $L_r$, including level adjustments instead of ASEL values alone.

Marshall and Davies investigated the perceptual attributes of recorded low amplitude booms and recordings of other transient environmental sounds. They found two major factors, the first one strongly related to loudness and startling and contributing to annoyance and the second one covering temporal properties and duration of the stimulus. A third factor was related to spectral balance. The second and the third factors were found to be especially important for sounds presented in a sonic boom simulator, accurately reproducing low frequency content, compared to experiments with earphones and high-pass filtered sounds. Due to the rather low levels and the different character of more recent low boom simulations, the evoked associations by these “new” sounds are not yet completely known.

The main objective of the present study was a comparison of ratings for low boom designs and conventional/N-wave signatures...
designs in a confined range of ASEL values around 60 dB(A) to see whether low boom signatures have a general benefit compared to conventional/N-wave signatures at similar ASEL values. Eight sonic boom signatures, including recent low boom simulations, conventional simulations, and recordings, were rated in terms of the perceived loudness and short-term annoyance. An indoor simulator at the University of Oldenburg was used to accurately reproduce the boom signatures for listening tests with voluntary participants. The simulator has been built for the assessment of human responses to low sonic boom signatures in the framework of the EU-project RUMBLE. Each signature was presented at three different signal levels to achieve a broad coverage of maximum overpressure values around 20 Pa and ASEL values around 60 dB(A). The relationships between listening test results and different single level-based metrics, A-, B-, C-, and D-weighted sound exposure levels (ASEL, BSEL, CSEL, and DSEL), were examined.

Another goal of this study was to investigate the influence of knowledge about the nature of the potential sound sources on the loudness and short-term annoyance ratings for the sonic boom sounds. Providing background information about prospective sound sources may affect the judgments of sounds in laboratory listening experiments. Therefore, information about the background of the study was presented only after completing a first rating session, which allows us to compare ratings as well as the evoked associations with and without information.

II. METHODS

A. Simulator setup

The listening tests took place in an indoor simulator that was constructed in a small room acting as a pressure chamber similar to facilities at NASA and at Japan Aerospace Exploration Agency (JAXA). The simulation room has a width of 1.15 m and enlarges to 1.44 m in the entrance area. The room is 2.60 m deep and has a tilted ceiling with an average height of 2.80 m. The volume of the pressure chamber is about 9 m³. To reduce the influence of the room acoustics and lower the reverberation time, especially for higher frequencies (above 300 Hz), absorbing material was placed on the walls and the ceiling of the pressure chamber. The walls to the left and the right of the listening positions were covered with pyramid foam absorbers (10 cm thickness). The wall at the end of the room (right side in Fig. 1) has been outfitted with soft foam loops (25 cm depth) to absorb low-middle frequencies. Figure 1 shows the inside of the indoor simulator at the University of Oldenburg. A neighboring room acts as a loudspeaker enclosure for two 18 in. speakers (2240 H, JBL, Northridge, CA) that drive the pressure chamber. This neighboring room is about 1.88 m wide, 2.75 m deep, and 3 m high with a slightly tilted ceiling. The loudspeaker chassis are mounted in a thick particle board installed in the door frame between the two rooms. The pressure chamber is accessed by another door with an airtight seal to the outside. The technical systems used to drive the low boom simulator are placed in a separate control room that is located one floor below the pressure chamber.

The two loudspeakers are driven in series by one power amplifier (BAA 500, Tira, Schalkau, Germany). The direct current (dc)-coupled input of the amplifier is connected to the dc-coupled headphone output of an audio interface (Fireface UFX+, RME, Haimhausen, Germany). Thus, the electric playback chain is capable of reproducing signals down to very low frequencies. Recordings in the chamber were made with a 1/2 in. infra sound microphone (47 AC, GRAS, Holte, Denmark) at the listening positions. The microphone is connected to a custom made integrated electronics piezo electric (IEPE) supply and amplifier that is linked to an analog-to-digital (AD) converter (ADI-8 QS, RME). The AD converter is digitally connected to the audio interface over an optical multichannel audio digital interface (MADI) cable. The electro-acoustic playback system is equalized with a digital filter for an accurate reproduction of the time signals at the listening position based on exponential sweep measurements in MATLAB (MathWorks, Natick, MA). Further details of the indoor simulator can be found in another publication.

The average background noise level in the simulator room due to noise from the outside is $L_{Aeq} = 21$ dB(A). Occasional background sounds from the outside, neighboring laboratories, and corridors can be heard and are not completely isolated from the simulator room. The reverberation time of the chamber averaged over octave bands from 63 Hz to 8 kHz is $T_{20} = 0.2$ s. Both values are comparable to the sonic boom simulators at NASA.

B. Stimuli

Different simulations and recordings of outdoor ground signatures from smaller and larger aircraft were used as stimuli in the listening tests. The signatures, shown in Fig. 2, include different low boom simulations (a)–(c), a simulated sharp (d) and soft (e) N-wave, and two recordings (f) and (g) from conventional supersonic aircraft. The signatures (a)–(g) originate from American Institute of Aeronautics and Astronautics (AIAA) sonic boom prediction workshops. The middle curve for signature (h) in Fig. 2 is a recent simulation of a low boom design (NASA, C25D). All original signatures were delivered by Sorbonne University in the context of the RUMBLE project.
Some of the signatures with N-wave shapes or an over-pressure substantially above 20 Pa were considerably attenuated to move them into a similar ASEL range as the low boom simulation C25D. In this way, all sounds could be accurately reproduced by the simulator, and excessive levels for the listeners were avoided. As shown in Fig. 2, each signature was varied in level in 3 dB steps over a 6 dB range. Thus, a broad coverage of ASEL values from 55.5 to 69.8 dB(A) and maximum overpressure values from 4.25 to 28.63 Pa was achieved (cf. Fig. 3). Table I summarizes the values of the ASEL and CSEL and the maximum overpressure ($p_{\text{max}}$) for the 24 stimuli, and Fig. 4 shows the spectrum of the medium level for each time signature.

C. Experimental procedure

The sonic boom signals were rated with respect to the loudness and the short-term annoyance by single listeners in listening tests. The participants were asked to give their annoyance ratings on an 11-point categorical scale. The categories had numerical labels from 0 to 10, and the ends of the scales had additional German verbal labels “überhaupt nicht lästig” (0, “not at all annoying”) and “extrem lästig” (10, “extremely

![FIG. 2. Time signals of the pressure for the eight different boom signatures used as input to the simulator. Each panel contains three curves for three different levels. Signatures (d)–(g) are N-waves and conventional booms, and signatures (a)–(c) and (h) are low boom simulations.](https://doi.org/10.1121/10.0003779)

![FIG. 3. Values of maximum overpressure ($p_{\text{max}}$) and ASEL for the eight signatures (a)–(h), indicated by different symbols, at three different levels each. Open symbols in (d)–(g) indicate N-waves and conventional boom signatures. Filled symbols in gray and black are low boom simulations.](https://doi.org/10.1121/10.0003779)

| Signature ID | ASEL [dB(A)] | CSEL [dB(C)] | $p_{\text{max}}$ (Pa) |
|--------------|--------------|--------------|-----------------------|
| (a)          | 67.5         | 91.8         | 17.01                 |
|              | 64.5         | 88.8         | 12.04                 |
|              | 61.5         | 85.8         | 8.52                  |
| (b)          | 61.5         | 91.7         | 26.56                 |
|              | 58.5         | 88.7         | 18.80                 |
|              | 55.5         | 85.7         | 13.31                 |
| (c)          | 64.0         | 89.7         | 24.53                 |
|              | 61.0         | 86.7         | 17.37                 |
|              | 58.0         | 83.7         | 12.30                 |
| (d)          | 69.8         | 88.2         | 8.49                  |
|              | 66.8         | 85.2         | 6.01                  |
|              | 63.8         | 82.2         | 4.25                  |
| (e)          | 66.2         | 97.4         | 28.63                 |
|              | 63.2         | 94.4         | 20.27                 |
|              | 60.2         | 91.4         | 14.35                 |
| (f)          | 69.1         | 95.9         | 25.01                 |
|              | 66.1         | 92.9         | 17.71                 |
|              | 63.1         | 89.9         | 12.54                 |
| (g)          | 57.9         | 88.9         | 28.55                 |
|              | 54.9         | 85.9         | 20.21                 |
|              | 51.9         | 82.9         | 14.31                 |
| (h)          | 64.5         | 94.1         | 27.41                 |
|              | 61.5         | 91.1         | 19.41                 |
|              | 58.5         | 88.1         | 13.73                 |
annoying”). The verbal scale labels were chosen to be simi-
lar to laboratory studies investigating the annoyance of
impulsive sounds of firearms19 and tramway noise32 as well
as indoor rattle noise15 and chair vibration4 in combination
with sonic booms heard indoors. A similar scale was also
proposed as one possibility to quantify annoyance in field
studies.33 However, the short-term annoyance measured in
laboratory experiments does not comprehensively reflect the
annoyance effect as experienced in real life but is rather
linked to the perceived unpleasantness of the sounds.34,35
Especially, the aspect of disturbance and the feeling of help-
lessness, which are both issues in noise annoyance,36 are
often not fully covered by annoyance ratings in laboratory
experiments. Therefore, we will use the term short-term
annoyance instead of annoyance in the following to clearly
distinguish the results of the present laboratory tests from
field test data.

In addition to the ratings of the short-term annoyance,
loudness ratings were collected in a separate task. For the

FIG. 4. Power spectra for the eight different boom signatures used as input
to the simulator. Each subfigure shows the unweighted (black line) and the
A-weighted power spectrum (gray line) for the medium level of each time
signature.

loudness task, the participants were asked, “How loud was
the sound on a scale from 0 to 10?” The loudness ratings
were given on an 11-point categorical scale from 0
(“überhaupt nicht laut,” “not loud at all”) to 10 (“extrem
laut,” “extremely loud”). The labels at the ends of the scale
were similar to the annoyance scale and laboratory studies
of effects of sonic boom shaping.10 The original questions
and scale labels were in German language for both tasks,
since all participants were native German speakers. The pre-
sentation of the stimuli and the graphical user interface for
the collection of the responses were implemented in MATLAB
(MathWorks).

A listening session started by giving the participant gen-
eral information about the listening experiments and collect-
ing written informed consent. Then the first listening
experiment, either the loudness or the short-term annoyance
task, took place in the simulator. Each experiment started
with written instructions and an orientation phase. In the ori-
entation phase, each of the 24 signals was played back to
give the participants a complete overview of the stimuli.
Then the first experiment started, and all 24 signals were
rated by the listener in a random order. Directly after each
experiment, the participant took a short break with some
questions from the investigator. The first open question
asked for the first impressions after the experiment, and the
second open question asked for associations with the sounds
and identified sound sources. After a short break, the second
experiment started with the orientation phase. One half of
the participants did the loudness task first and the short-term
annoyance task second in each session. The other half car-
rried out the two tasks in reverse order. The ratio of female
and male participants was balanced over the two groups of
participants.

The listening tests were divided into two sessions on
different days. In the first session, the participants received
as little information about the background of the study as
possible prior to the stimulus presentations. Only after hav-
ing finished both listening tests of the first session was each
participant debriefed and finally informed about the back-
ground of the study. The participants were informed that the
study was concerned with the perception of prospective low
sonic boom sounds. They were further briefed that the
sounds originated from supersonic aircraft and that civil
supersonic flights were allowed over water only to avoid
detrimental effects for inhabitants of overflown areas.

The second session was a repetition of the same listen-
ing experiments. The same information that was used for the
debriefing at the end of the first session was given at the start
of the second session to ensure that all participants had the
same amount of prior knowledge. Between the first and the
second session was a gap of at least 2 days (one weekend)
and in most cases about one week. The order of the two
tasks was kept the same as in the first session for each par-
ticipant, but the order of the signals was newly randomized
for each task. At the end of the second session, the partici-
pants had to fill out a questionnaire pertaining to noise sensi-
tivity (NoiSeQ-R37,38) and attitude toward traffic and traffic
noise as well as satisfaction with their current living environment.

D. Participants

A total of 16 volunteers (10 female, 6 male) participated in the listening tests. The participants had an average age of 24 yrs (age range from 19 to 29 yrs). About 56% of the participants (eight female, one male) had prior experience with other listening experiments. The other 44% had no prior experience with listening tests. All of the participants declared that they had no hearing problems.

After the end of the listening tests, the participants were asked whether they had heard a real sonic boom before taking part in the study. Nine participants (seven female, two male) had heard a sonic boom before. The rest of the participants had either not heard a sonic boom before (two female, two male) or were not sure (one female, two male).

Each participant was paid a compensation of €20 for the two sessions (€10 per hour). The commission for research impact assessment and ethics of the University of Oldenburg had no objections regarding the listening experiments of this study (ethics application EK/2018/104).

III. RESULTS

The results of a four-way analysis of variance (ANOVA) with repeated measures and two between-subject factors, calculated with SPSS 25 (IBM, Armonk, NY), are presented in Sec. III A and the following. Within-subject factors were signature [(a)–(h)] and relative level (−3, 0, or +3 dB), session [1 (without information) or 2 (with information)], and task (loudness or short-term annoyance). Mauchly’s test indicated that the assumption of sphericity had not been violated for any of the within-subject factors. Between-subject factors were the order of tasks (annoyance → loudness or loudness → annoyance) and the gender (female or male) presented in Sec. III E. All effects are reported as significant at $p < 0.05$. In Secs. III F and III G, qualitative results from the interviews after the listening experiments are presented.

A. Influence of time signatures and level changes

Figure 5 shows the average short-term annoyance and the loudness ratings for the eight different signatures at three different levels each. The average ratings cover a broad range of scale values between 1 and 8.5 scale units for short-term annoyance ratings and for the loudness ratings. In general, the ratings differ between the signatures, with a significant main effect for signature $F(7, 84) = 153.20$, $p < 0.05$, $\eta_p^2 = 0.93$, and increase with rising levels indicated by a main effect of the relative level, $F(2, 24) = 436.35$, $p < 0.05$, $\eta_p^2 = 0.97$. There was also a significant interaction effect for signature $\times$ level, $F(14, 168) = 4.34$, $p < 0.05$, $\eta_p^2 = 0.27$. This means that the variation of the relative level affected the judgments differently for different signatures. The delta in short-term annoyance ratings resulting from the 6 dB variation was larger for some signatures [e.g., (e) and (f)] compared to others [e.g., (h)].

Furthermore, there were significant interaction effects for signature $\times$ order, $F(7, 84) = 3.39$, $p < 0.05$, $\eta_p^2 = 0.22$; a significant interaction effect for signature $\times$ gender, $F(7, 84) = 5.66$, $p < 0.05$, $\eta_p^2 = 0.32$; and a significant interaction effect for signature $\times$ order $\times$ gender, $F(7, 84) = 2.26$, $p < 0.05$, $\eta_p^2 = 0.16$. Similarly, there were significant interaction effects for level $\times$ order, $F(2, 24) = 7.77$, $p < 0.05$, $\eta_p^2 = 0.39$; for level $\times$ gender, $F(2, 24) = 6.37$, $p < 0.05$, $\eta_p^2 = 0.35$; and for level $\times$ order $\times$ gender, $F(2, 24) = 9.43$, $p < 0.05$, $\eta_p^2 = 0.44$. The 95% confidence intervals shown as error bars in Fig. 5 are smaller than 0.8 scale units, indicating a rather high agreement among the participants. The effects of the between-subject factors order and gender are further presented and discussed in Sec. III E.

B. Relationship between average ratings and acoustic metrics

The results of the short-term annoyance ratings from the two sessions averaged across all 16 participants are plotted over the ASEL values of the signals in Fig. 6. The average annoyance ratings increase with rising ASEL values for each of the signatures, and the correlation coefficient (Pearson’s $r$) between ASEL values and the average short-term annoyance ratings for all sounds was statistically
significant ($\rho = 0.940, p < 0.001$). Signature (f) results in the most annoying signals, while signature (a) results in the least annoying signals compared to other signals in a similar range of ASEL values. The signal with the lowest ASEL from signature (g) is the overall least annoying one. Signature (h) and the other low boom simulations (b) and (c) were rated similar in terms of short-term annoyance as the N-waves and the recorded conventional boom sounds when having the similar ASEL values. Please note that some signatures, like the signals for (d), (f), and (g), were considerably attenuated compared to their original level to move them to a common level range with the low boom simulations. These signatures would probably occur at higher levels in reality and be considerably louder and more annoying than shown in the present results.

Correlation coefficients between different commonly used level-based metrics and the average ratings for short-term annoyance and loudness are given in Table II, and the correlations with the short-term annoyance ratings are shown in Figs. 6 and 7. Due to the systematic variation of signature levels, the predictive capabilities of level metrics are considered here. Statistically highly significant correlation coefficients ($p < 0.001$) were found for ASEL, BSEL, and DSEL and $p = 0.002$ for CSEL. Among the different sound exposure level metrics, the highest correlation coefficient was found for ASEL. This finding suggests that middle and high frequencies above 1 kHz are more important for the loudness and annoyance ratings than the amount of low-middle (below 1 kHz) and low frequencies. There is a remaining variability in the ratings of about 2 scale units between sounds with similar ASEL values, which might be related to differences between signatures in perceived sound character. Taking the results in Fig. 5 into account, the relative differences of about 2 scale units are equal to a level difference between 3 and 6 dB, depending on the signature. For all other sound exposure level (SEL) metrics, the variability in short-term annoyance ratings for sounds having similar metric values is larger than for ASEL, resulting in lower correlation coefficients for these metrics. The largest variability in the ratings is observed for CSEL, with a variability of about 5 scale units in the middle panel of Fig. 7 while the variability for ASEL is only about 2 scale units in Fig. 6. Very low frequencies, which are higher weighted by the C-weighting than by the A-weighting, are apparently not the driving factor for the annoyance of sonic boom at low levels, but the middle and high frequencies are. A systematic

![Graph](image-url)
benefit of low boom signals compared to conventional/N-wave signatures having a similar weighted SEL was not observed, and the open and filled markers are intermixed for all four SEL metrics (cf. Figs. 6 and 7).

C. Influence of information about the study’s background on the loudness and annoyance ratings

The main effect of the factor session was not statistically significant, $F(1, 12) = 0.33, p = 0.58, \eta^2_p = 0.03$. Figure 8 shows the relationship between the average ratings for the first session (session 1) without information and the second session (session 2) with information in separate subplots for the loudness and the annoyance ratings, respectively. The average ratings from the second session are highly correlated with those from the first session, which is in line with the insignificant main effect of the factor session in the ANOVA results. Pearson’s $r$ was statistically significant in the case of the loudness task ($r = 0.988, p < 0.0001$) and for the annoyance task ($r = 0.986, p < 0.0001$). Surprisingly, the information on the background of the study and prospective future sound sources seems to have no considerable influence on the judgments in the listening tests. Only for some of the extremely annoying sounds, the average short-term annoyance ratings of the second session are slightly higher compared to those of the first session.

D. Relationships between loudness and short-term annoyance ratings

The main effect of the factor task was not statistically significant, $F(1, 12) = 1.01, p = 0.34, \eta^2_p = 0.08$, but there was a significant interaction effect for task $\times$ order $\times$ gender, $F(1, 12) = 11.12, p < 0.05, \eta^2_p = 0.48$. Figure 9 shows the relationship between the average loudness and short-term annoyance ratings in separate subplots for the first session (without information) and the second session (with information). The average short-term annoyance ratings are tightly linked to the loudness judgments, and the average values for short-term annoyance and loudness are nearly identical (on the diagonal line in Fig. 9). In the results from the first session (Fig. 9, left), only some of the average short-term annoyance values are slightly lower than the corresponding loudness values. In line with the insignificant main effect of the factor task in the ANOVA results, the average ratings from the short-term annoyance task are highly correlated with those of the loudness task. This is the case for the first session without background information ($r = 0.990, p < 0.0001$) and for the second session with background information ($r = 0.994, p < 0.0001$). The average short-term annoyance judgments are thus tightly linked to the loudness sensation in each of the two sessions of this study.

E. Influence of task order and gender

For the between-subject effects, there was a significant effect for the order of the tasks, $F(1, 12) = 5.22, p < 0.05, \eta^2_p = 0.30$. On average, the ratings were 0.62 scale units higher if the annoyance task was first and the loudness task second compared to the opposite order of the tasks. Figure 10 shows that this effect is mainly driven by the signatures (a), (d), (e), and (f).

There was also a significant effect of gender, $F(1, 12) = 28.44, p < 0.05, \eta^2_p = 0.70$. On average, female participants rated the low boom sounds 1.44 scale units

FIG. 8. Average ratings ($N = 16$) from session 2 (with information) plotted over the results from session 1 (without information) for the loudness ratings (left) and short-term annoyance (right).

FIG. 9. Relationship between the average loudness and short-term annoyance ratings for all 24 stimuli. Shown are average values ($N = 16$) from the first session (left) without background information and the second session (right) with information on the background of the study. The dashed line indicates identical ratings.

FIG. 10. Ratings averaged across sessions, tasks, and participants ($N = 8$) for the factor order of the two tasks. The symbols indicating the eight different signatures are slightly shifted on the abscissa to visually disentangle the error bars that show the 95% confidence interval.
higher (more annoying and louder) than the male participants. Figure 11 shows that the difference between the ratings coming from female and male participants is largest for the signatures (c), (d), and (f). Note that although there is a significant effect of gender, it leads to a change in ranking order for only two sounds in our current experiment (cf. Fig. 11). Eventually, there was a significant interaction effect of order for only two sounds in our current experiment (cf. Fig. 11).

The results from the NoiSeQ-R questionnaire are given in Table III. In total, the participants have an average noise sensitivity of 1.64 (±0.60) on the NoiSeQ-R scale from 0 to 4. The group of female participants has a slightly higher noise sensitivity score of 1.79 (±0.58) compared to 1.39 (±0.58) for the male group. The higher noise sensitivity score of the female participants (see Table III) might be a factor for explaining the higher short-term annoyance and loudness ratings compared to the male participants.

F. Statements of the participants in the interviews after the listening tests

Directly after each listening experiment, the participants were asked about their impressions and associations with the sounds they had heard. Several very different associations were mentioned by the participants. Figure 12 shows the number of participants who mentioned a certain item for the most frequently stated aspects.

In the first session, the participants did not receive any information about the background of the study and the potential source of the sounds. After the first listening test, half of the participants associated the sounds with knocking on a door or a wall. Six participants associated sounds from music, bass, drum, music over headphones, concerts, bass in electronic music, or bass tones. Five participants had associations with something falling over, falling down, or being moved. A quarter of the participants mentioned heartbeat or a heart beating as an association. Apparently, the presented sounds were mainly associated with neutral or even positively connoted items and not linked to immediate fear or direct threat. Only one male participant of the overall 16 participants mentioned an aircraft breaking the sound barrier as a sound source.

After the second listening experiment of the first session to make sure every participant had the information on the background of the study. At the end of the first session of listening experiments, all participants were debriefed and informed about the background of the study and future supersonic aircraft as prospective sound sources. The information for the debriefing was used again at the start of the second session to make sure every participant had the information on the background of the study.

After the first experiment of the second session, which took place on a different day, many aspects were described similarly like in the first session. However, the aspect of something falling over down was mentioned only a single time after the first experiment of the second session. Four participants stated after the first experiment that, in accordance with the information on the study given to them, they associated aircraft sounds and sonic booms with the presented sounds. Six participants mentioned that the sounds were not an aircraft, not coming from an aircraft, or not

![FIG. 11. Ratings averaged across sessions, tasks, and levels for the factor gender. The symbols indicating the eight different signatures are slightly shifted on the abscissa to disentangle the error bars that show the 95% confidence interval.](https://doi.org/10.1121/10.0003779)

**TABLE III.** Noise sensitivity scores derived from the results of the NoiSeQ-R questionnaire. Given are the average values and standard deviations for the three subscales (sleep, habitation, and work) and the total values for the female participants, male participants, and all participants.

| Subscale | Female (N = 10) | Male (N = 6) | All (N = 16) |
|----------|----------------|-------------|-------------|
| Sleep    | 1.78 ± 0.42    | 1.21 ± 0.43 | 1.56 ± 0.50 |
| Habitation | 1.60 ± 0.47    | 1.33 ± 0.70 | 1.50 ± 0.61 |
| Work     | 2.00 ± 0.44    | 1.63 ± 0.59 | 1.86 ± 0.66 |
| Total    | 1.79 ± 0.58    | 1.39 ± 0.58 | 1.64 ± 0.60 |

![FIG. 12. Counts of the most frequently stated items (in italics) from session 1, without background information (left), and session 2, with background information (right). Filled bars show the count of mentions after the first experiment and open bars after the second experiment of a session. Further items stated only once are not shown here but mentioned in the text.](https://doi.org/10.1121/10.0003779)
being a sonic boom. Of the six participants, four female participants had heard a sonic boom before their participation in the current study began, while the two male participants had not heard a sonic boom before or were unsure about that. This explicit statement of the participants might be interpreted as an indication for a lack of realism in the listening tests. It is also possible that the sounds were too unfamiliar and too unknown to the participants to be associated with an aircraft. The aspect (not an aircraft) was not stated again after the second experiment of this session. Despite the given information about the study’s background, 6 of the 16 participants did not make any references to aircraft or sonic boom sounds in their associations and potential sound sources at all.

Beyond the results shown in Fig. 12, there were some further associations stated by individual participants, including a subway or a tram, a slamming door, pressure compensation, an explosion when wearing a headphone, a ball bouncing, somebody stomping on the floor, pot-banging, somebody running down stairs, a firecracker thrown into a manhole, and a jackhammer at a construction site. Overall, these qualitative results indicate that it can be difficult to convince participants of a certain noise source in a laboratory listening test. The rather low sound pressure levels for some of the signals might have contributed to this circumstance, especially in the context of sonic booms.

G. Criteria for annoyance judgments

At the end of each session, the participants were asked for the acoustics aspects underlying the annoyance judgments. Nearly all participants (15 of 16 participants) mentioned the loudness of the sounds as a criterion. This statement is not unexpected, given the within-subject design with a loudness and a short-term annoyance task, and it is manifested by the tight link between the ratings from both tasks (cf. Sec. III D).

However, other acoustic aspects were mentioned as well. Clear or shrill sounds were perceived as more annoying across all presentation levels by four participants in session 1 and one in session 2, while one participant judged sounds that were not clear as more annoying. In both sessions, two participants found dull/low sounds more annoying, while for three participants dull sounds were less annoying/more pleasant. Sounds with a distinct over- and underpressure sounded like two bangs directly after each other, whereas signature (a) sounded more like a single low frequency scrunch. Five participants found two sounds perceived quickly after each other or a knocking sound more annoying in both sessions, while one participant stated that a single bang was more annoying than the knocking sound in session 2. Some participants mentioned that the time between the two sounds did not play a role for them (two participants in session 1 and one in session 2). Further acoustic aspects linked to the annoyance were when “the sounds could be felt” or if “it included such a pressure.” These statements indicate the importance of very low frequency content for the annoyance ratings, which is in contrast to the rather low correlation coefficients for metrics giving more weight to low frequency content (e.g., CSEL) than ASEL in Sec. III B.

IV. DISCUSSION

A. Relationship between average ratings and acoustic metrics

The different ratings between signatures with similar maximum overpressure but different signature shapes is in general agreement with the results of Leatherwood and Sullivan. In their study, signatures with a steep front-shock rise time were louder than those with a shallower front-shock. In the present study, N-wave signatures, like signatures (d) and (e) in Fig. 2, are rated to be louder and also more annoying than signatures with a shallower rise, like signature (g) in Fig. 2, given similar maximum overpressure values.

The correlation coefficients between the average subjective ratings and the ASEL values from the present study (cf. Table II) are similar to those found in literature. Leatherwood and Sullivan found correlation coefficients around 0.96 between ASEL values and loudness ratings for simulations created from connected straight lines. Apparently, the loudness and annoyance differences between simple simulations like N-waves, created from straight lines, and more complex shapes is reflected quite well by ASEL values. This was also found by Sullivan. The correlation coefficients in the present study are higher for ASEL than for CSEL, which is in overall agreement with results found in literature. The present correlation coefficients are also higher for BSEL and DSEL compared to CSEL as in Refs. 3, 9, and 41. The effect of different frequency weightings on the correlation coefficients is similar to studies that investigated a similar dynamic range of ASEL values and similar or even more pronounced than in studies that investigated a larger dynamic range of ASEL values.

Vos investigated the annoyance produced by a broad variety of small, medium-large, and large firearms. He achieved a more precise and unified characterization of the annoyance from single impulsive events by fitting a rating sound level \( L_r \), including level adjustments to his data instead of using ASEL values alone. The rating level \( L_r \) was defined by Vos as

\[
L_r = L_{AE} + 12 \text{ dB} + \beta \cdot (L_{CE} - L_{AE}) \cdot (L_{AE} - \alpha) \text{ dB},
\]

in which \( \alpha = 45 \text{ dB} \) and \( \beta = 0.015 \text{ dB}^{-1} \). The correlation coefficient between \( L_r \) values and the annoyance ratings was statistically significant (\( p = 0.943, p < 0.001 \)) for the

\[\text{https://doi.org/10.1121/10.0003779} \]
present data and only slightly higher than that for ASEL. Thus, the rating level proposed by Vos for impulsive sounds from firearms is, if at all, only marginally better suited for the description of the low sonic booms investigated in this study than the ASEL alone.

Other studies indicate that psychoacoustic loudness models, or the inclusion of a loudness derivative and duration for the description of the sound character in multi-metric annoyance models can enable a better characterization of the perception of transient sounds than single level-based measures. A description of the present ratings with loudness and other psychoacoustic models might be worthwhile in the further analysis of the present results.

**B. Link between loudness and annoyance judgments**

In the present study, a tight link between loudness and short-term annoyance ratings was observed for low sonic boom sounds. Similarly, Marshall and Davies found high loadings of the adjective pairs “not annoying–annoying” and “soft–loud” on a common factor, investigating low amplitude sonic boom and other transient sounds. Such a close correlation between loudness and annoyance judgments is often found in laboratory listening tests, whereas annoyance measured in the field includes further factors like the cognitive image, avoidability, and disturbance.

However, the order of the loudness and the short-term annoyance task had a statistically significant influence on the ratings in the present study. This indicates that although the two aspects are tightly linked to each other, they were apparently not understood as interchangeable by the participants. Using a different experimental method, it might be possible to further distinguish between the two aspects. In other studies, it was already shown that a clear distinction between loudness and preference judgments is possible, especially for unpleasant sounds, by measuring points of subjective equality with an adaptive variation of the overall level. By applying that measurement method, the estimated equivalent of the remaining variability equivalent to 3–6 dB might be confirmed.

**C. Criteria applied by the participants**

In the present study, five participants clearly stated that two bangs occurring quickly after each other were more annoying for them, which is in contrast to the observation that Fidell et al. made. They did not observe a distinction between annoyance judgments of shorter- or longer-duration simulated sonic booms. In the present study, only a few participants mentioned that the time between the two bangs did not play a role for them. Another conclusion of Fidell et al., that the low frequency content may control the annoyance, is supported by some of the participants of the present study. On the other hand, other participants found clear and shrill sounds more annoying or dull and low sounds less annoying or more pleasant. The broad variety of different signatures in a similar level range might have contributed to the diversity in judgment criteria of the present study.

Marshall and Davies explored the perceptual dimensions of low sonic booms and other transient environmental sounds. The three most important factors in their study were related to (1) loudness, startle, and contribution to annoyance; (2) temporal factors and duration of the signals; and (3) spectral balance of the signals. The second and third factors were found to be more important for a sonic boom simulator experiment than in an experiment with earphones and high-pass filtered sounds. The criteria for the annoyance judgments stated by the participants in the present study (in Sec. IV C) included (1) how loud, (2) how quickly the bangs occurred one after another, and (3) how dull/low or shrill/clear the sounds were. Thus, the perceptual aspects mentioned by the participants in the present study confirm the perceptual factors identified by Marshall and Davies for the simulator experiments in general.

**V. CONCLUSION**

Loudness and short-term annoyance ratings were collected for conventional and low boom signatures at low overall levels. The main effects for the factors signature and relative level as well as an interaction effect between the two were statistically significant. The average annoyance ratings were best reflected by ASEL values compared to other frequency weightings. The results of the listening experiments show that the tested low boom designs are similarly loud and annoying as conventional and N-wave sounds when presented at similar ASEL values, and an indication of a systematic benefit for low boom signals compared to conventional booms was not observed. The observed remaining variability of about 2 scale units between sounds with similar ASEL values may be estimated as a level equivalent of 3–6 dB based on the link between relative level changes and short-term annoyance ratings.

Surprisingly, information on the background of the study and the prospective sources of sounds did not significantly affect the ratings of sonic boom sounds in the present laboratory study. The average ratings of two successive sessions were highly correlated to each other for the loudness as well as the short-term annoyance task. This finding suggests that the participants primarily focused on the acoustic stimuli for a rather direct formation of their judgments in the artificial laboratory environment. Further evidence for a direct focus on the stimuli is the observed tight link between loudness and short-term annoyance ratings, although the two concepts were not understood as completely interchangeable by the participants, shown by a significant task order effect.

Future studies could include imagined or simulated contexts to obtain ratings that are closer to a prospective real life assessment and more ecologically valid. Potentially, including room acoustic reverberation might also contribute to the perceived realism and mightiness by giving listeners an impression of a large event affecting the surrounding building even if a sound is not perceived as very loud. Including post-boom rumble could further improve the
realism and could be more important than the reproduction of the very low frequency content.  

ACKNOWLEDGMENTS

The RUMBLE project has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement No. 769896. We would like to thank our colleagues in RUMBLE for the exchange during the preparation of this study and one anonymous reviewer for the detailed remarks and very helpful suggestions.

1V. W. Sparrow, “Lowering the boom,” Acoust. Today 2(1), 20–28 (2006).
2D. J. Maglieri, P. J. Bobbittta, K. J. Plotkin, K. P. Shepherd, and P. G. Coen, “Sonic Boom: Six Decades of Research,” NASA/SP-2014-622, NASA Langley Research (2014).
3D. Carr, P. Davies, A. Loubeau, J. Rathsam, and J. Klos, “Influences of low-frequency energy and testing environment on annoyance responses to supersonic aircraft noise when heard indoors,” J. Acoust. Soc. Am. 148(1), 414–429 (2020).
4J. Rathsam, J. Klos, A. Loubeau, D. J. Carr, and P. Davies, “Effects of chair vibration on indoor annoyance ratings of sonic booms,” J. Acoust. Soc. Am. 143(1), 489–499 (2018).
5A. Loubeau and J. Page, “Human perception of sonic booms from supersonic aircraft,” Acoust. Today 14(3), 25–30 (2018).
6Y. Makino and Y. Naka, “Sonic-boom research and low-boom demonstrator projects in JAXA,” in Proceedings of ICA 2007, Madrid, Spain (September 2–7, 2007), pp. 1–6.
7F. Coulouvrat, “The challenges of defining an acceptable sonic boom overland,” in Proceedings of the 15th AIAA/CEAS Aeracooustics Conference (30th AIAA Aeracooustics Conference), pp. 1–13 (2009).
8J. D. Leatherwood, K. P. Shepherd, and B. M. Sullivan, “A new simulator for assessing subjective effects of sonic booms,” Technical Memorandum 104150, NASA Langley Research (1991).
9Y. Naka, “Subjective evaluation of loudness of sonic booms indoors and outdoors,” Acoust. Sci. Technol. 34(3), 225–228 (2013).
10J. D. Leatherwood and B. M. Sullivan, “Laboratory study of effects of sonic boom shaping on subjective loudness and acceptability,” NASA Technical Paper NASA-TP-3269, NASA Langley Research (1992).
11J. D. Leatherwood and B. M. Sullivan, “Loudness and annoyance response to simulated outdoor and indoor sonic boom,” NASA Technical Memorandum NASA-TM-107756, NASA Langley Research (1993).
12J. D. Leatherwood and B. M. Sullivan, “A laboratory study of subjective annoyance response to sonic booms and aircraft flyovers,” NASA Technical Memorandum NASA-TM-109113, NASA Langley Research (1994).
13D. Fidell, L. Silvati, and K. Pearsons, “Relative rates of growth of annoyance of impulsive and non-impulsive noises,” J. Acoust. Soc. Am. 111(1), 576–585 (2002).
14G. M. Sullivan, J. Klos, R. D. Buehrle, D. A. McCurdy, and J. E. A. Haering, “Human response to low-intensity sonic booms heard indoors and outdoors,” NASA Technical Memorandum NASA/TM-2010-216685, NASA Langley Research (2010).
15J. Rathsam, A. Loubeau, and J. Klos, “Effects of indoor rattle sounds on annoyance caused by sonic booms,” J. Acoust. Soc. Am. 138(1), EL43–EL48 (2015).
16B. M. Sullivan, P. Davies, K. K. Hodgdon, J. A. Salamone, and A. Pilon, “Realism assessment of sonic boom simulators,” Noise Control Eng. J. 56(2), 141–157 (2008).
17B. M. Sullivan, “Human response to simulated low-intensity sonic booms,” in Proceedings of Noise-Con 2004, Baltimore, MD (July 12–14, 2004), pp. 1–10.
18J. Rathsam, A. Loubeau, and J. Klos, “A study in a new test facility on indoor annoyance caused by sonic booms,” NASA Technical Memorandum NASA/TM-2012-217332, NASA Langley Research (2012).
41A. Loubeau, Y. Naka, B. G. Cook, V. W. Sparrow, and J. M. Morgenstern, “A new evaluation of noise metrics for sonic booms using existing data,” AIP Conf. Proc. 1685, 090015 (2015).
42S. Töpken, H. Scheel, J. L. Verhey, and R. Weber, “Quantification of preference relevant sound characteristics of multi-tone sounds based on the differences between loudness judgments and preference evaluations,” Acta Acust. united Ac. 104(1), 153–165 (2018).
43S. Töpken and S. van de Par, “Determination of preference-equivalent levels for fan noise and their prediction by indices based on specific loudness patterns,” J. Acoust. Soc. Am. 145(6), 3399–3409 (2019).
44C. Marquis-Favre, “Laboratory methods with imaginary and simulated contexts to assess noise annoyance: A comparison in terms of annoyance model testing,” Acta Acust. united Ac. 105(6), 1078–1085 (2019).