Short-term annoyance reactions to civil helicopter and propeller-driven aircraft noise: A laboratory experiment

Armin Taghipour,* Reto Pieren, and Beat Schäffer
Empa, Swiss Federal Laboratories for Materials Science and Technology, Laboratory for Acoustics/Noise Control, Überlandstrasse 129, 8600 Dübendorf, Switzerland

(Received 4 October 2018; revised 21 January 2019; accepted 28 January 2019; published online 21 February 2019)

Helicopter noise exhibits distinctive acoustical characteristics (e.g., pulsation) compared to noise from propeller-driven aircraft which contains tonal components. Whereas, at comparable sound exposure levels \( L_{AE} \), annoyance reactions to these sources might be different, knowledge of potential annoyance differences is scarce. This paper reports a comparison between short-term annoyance reactions to noise from light-weight helicopters and propeller-driven aircraft in a laboratory setup. Stimuli were presented with a 3D sound reproduction system in a listening test facility based on field recordings of takeoffs and landings. Propagation filtering and amplitude changes were carried out to simulate various propagation distances and source levels, covering a reasonable \( L_{AE} \) range from 64 to 85 dB(A) for a stimuli length of 24 s. Fifty-six subjects rated their short-term annoyance reactions on the ICBEN 11-point numerical scale. Associations between design variables (source type, procedure, and \( L_{AE} \)) and short-term annoyance were explored by means of a linear mixed-effect model. \( L_{AE} \) was found to be the major predictor. For the major range of \( L_{AE} \), no significant difference was found between annoyance to noise from the two aircraft types. Observed level differences at equal annoyance ratings were below 1 dB. Furthermore, helicopter landings were found slightly more annoying than helicopter takeoffs. © 2019 Acoustical Society of America. https://doi.org/10.1121/1.5090500

I. INTRODUCTION

Exposure to aircraft noise can be annoying to residents living close to an airfield and can have further health impacts (Guski et al., 2017; WHO, 2011, 2018). Noise emitted by helicopters, propeller-driven aircraft and passenger (jet) aircraft differ to some extent in their acoustical characteristics. Specifically, helicopters exhibit, among others, pulsating noise and rotor blade slaps (Leverton, 1975), while noise of propeller-driven aircraft often has tonal components. The distinctive sound characteristics of helicopters might be associated with different annoyance reactions as compared to other aircraft. Knowledge of these potential annoyance differences is, however, scarce. The present study aimed to investigate differences between annoyance reactions to noise from helicopters and propeller-driven aircraft in a laboratory setup, i.e., whether—at comparable A-weighted sound exposure levels \( L_{AE} \)—there would be any noise source type specific differences between annoyance reactions associated with these two types of aircraft.

Noise assessment regulation authorities in most countries (such as Germany, Netherlands, Norway, UK, USA, etc.) do not currently suggest any helicopter-specific correction while assessing the noise community impact. The current Swiss noise assessment regulation (LSV, 1986) requires, in the case of helicopter airfields, taking \( L_{AS\text{max}} \) (maximum A-weighted noise level) into account in addition to assessment of the common A-weighted equivalent sound level \( L_{A\text{eq}} \). It is questionable whether such aircraft type specific treatment is necessary, or whether the annoyance due to noise can be explained sufficiently by the \( L_{A\text{eq}} \) or, in case of individual flight events, the \( L_{AE} \).

Unfortunately, the relatively rich literature on helicopter noise is rather scarce and ambiguous in comparing annoyance reactions to helicopters to other aircraft types. Leverton (1975) suggested that conventional acoustical assessment methods used for traffic noise and aircraft might not be suitable for helicopters and additional corrections for impulsiveness might be needed.

Schomer (1983) conducted a community survey on aircraft noise. Annoyance to airplane noise correlated well with \( L_{dn} \) (day-night sound level) for all aircraft types, but not for helicopters. Respondents were more annoyed by helicopters than \( L_{dn} \) suggested. Also Miedema and Oudshoorn (2001) noted regarding their exposure-response relationships that they might be invalid for helicopters.

A review of 30 studies on helicopter noise, which investigated possible impulse corrections for helicopter noise based on crest level types of measurement, was carried out by Molino (1982). He reported that 18 studies suggested the need for a crest level based impulse correction (commonly by an increase of 2–3 dB) before assessing the community impact. However, whether a helicopter was more annoying than other aircraft types seemed to be depending on the type of the helicopter. The remaining 12 studies did not report any need for such level corrections. In line with this, Atkins et al. (1983) and Newman and Beattie (1985) recommended to use \( L_{dn} \) for all airfields, regardless of existence or absence of helicopter flights.

*Electronic mail: armin.taghipour@empa.ch
Fields and Powell (1987) studied community reaction to helicopter noise events in a field experiment. Annoyance increased with sound pressure level, as well as with logarithmic number of flights. That is, when few flights were present, an increment in the number of flights led to a considerable annoyance increase; however, in the presence of a large number of flights, additional flights did not contribute much to annoyance. Furthermore, a non-significant trend was found for increased annoyance due to impulsive helicopters. Contradicting the latter findings, Powell (1981) found that less-impulsive helicopters were more annoying. Accordingly, a number of studies concluded that the impulsiveness of helicopters should not be taken into account when assessing community impact of the noise (Berry et al., 1979; FAA, 2004).

Gjestland (1994) conducted a laboratory study comparing helicopter noise to jet aircraft noise.1 Depending on the type of helicopter and jet aircraft and on the procedure (takeoff vs landing), one or the other seemed to be subjectively more annoying. Whereas small and big helicopters were found to be least and most annoying, respectively, jet aircraft noise was rated in between. As the results were not clearly in one direction, Gjestland (1994) recommended to apply the same conventional assessment procedure for both fixed wing and rotary wing aircraft, without any correction factors. A recent extensive literature review and a community noise survey by Mestre et al. (2017) came to similar conclusions. No notable differences could be found between annoyance to light civil helicopter noise and the noise generated by fixed-wing aircraft at comparable sound exposure levels (Mestre et al., 2017).

Furthermore, literature shows that also non-acoustical attitude factors can influence (and in cases increase) the perceived annoyance by helicopters; e.g., fear of helicopter crash, attitude towards military helicopter operations, etc. (Atkins et al., 1983; Fields and Powell, 1987; Janssen et al., 2017; Ollerhead and Jones, 1994; Schomer, 1983). On the contrary, Mestre et al. (2017) reported that no non-acoustic factors could be identified that might affect an individual’s perception of helicopter noise.

The literature indicates no consistency/definiteness regarding annoyance reactions to helicopter noise (compared to other types of aircraft). The majority of comparisons was with military and passenger jet aircraft. Two types of results and recommendations were found: either no level corrections were suggested, or if any, corrections were recommended for helicopters. Most such level corrections were about 2–3 dB. No study was found suggesting a level correction in favor of helicopters.

In the literature mentioned above, helicopters were commonly compared to various jet aircraft types which were considerably heavier and larger than helicopters. With this background, the study reported in the present paper aimed to make a comparison between short-term annoyance reactions to light-weight civil helicopter noise and to noise of propeller-driven aircraft, which seem to be an appropriate counterpart to helicopters because both these aircraft types are frequently present in small airfields, exhibit comparable propagation distance to the observer, and are more comparable in size/weight.

II. METHOD

Acute annoyance reactions to helicopters and propeller-driven aircraft were investigated under laboratory conditions. The observed annoyance ratings correspond to “short-term annoyance” (Bolin et al., 2014) or “psychoacoustic annoyance” (Fastl and Zwicker, 2007).

A. Listening test concept

To investigate differences in short-term annoyance, stimuli were generated from field recordings of helicopters and propeller-driven aircraft during takeoff and landing. Three design parameters, i.e., two categorical variables source type (helicopter and propeller-driven aircraft) and procedure (takeoff and landing) as well as the continuous variable A-weighted sound exposure level, $L_{AE}$ (covering a reasonable and comparable level range) were systematically varied to study their individual and combined associations with short-term annoyance ratings.

B. Listening test facility

The experiment was conducted in the listening test facility “AuraLab” of Empa in Switzerland. AuraLab has a separate listening and control room allowing for audio-visual supervision to comply with ethical requirements.

AuraLab satisfies room acoustical requirements for high-quality audio reproduction in terms of its background noise [below 7 dB(A)] and reverberation time. For the present experiment, floor sound absorbers were additionally installed. The experiment setup in AuraLab is shown in Fig. 1. The subject is seated in the central listening spot.

During the experiment, a 3D immersive sound system with 16 separate audio channels was used. Fifteen loudspeakers “KH 120 A” (Georg Neumann GmbH, Berlin, Germany) were used to generate the stimuli. Each loudspeaker was connected to a digital audio workstation, and the audio signals were sent to a digital audio controller. The audio signals were then sent to a 16-channel power amplifier, and the amplified signals were sent to the loudspeakers.

FIG. 1. (Color online) Experimental setup in AuraLab with a loudspeaker array according to Fig. 2.
Germany) were located in a hemispherical arrangement on three height levels (0°, 30°, and 60° vertically) in a distance of 2 m from the central listening spot, as depicted in Fig. 2. Bass management by means of two subwoofers “KH 805” (Georg Neumann GmbH, Berlin, Germany) and a digital signal processor (DSP) completed the playback system.

C. Stimuli

A set of 46 stimuli was prepared for the listening test. Flight events (i.e., takeoffs and landings of a number of different helicopters and propeller-driven aircraft) were mixed with low-level ambient sounds of bird and vegetation. Their recording, processing and playback, and a pilot experiment on their duration will be discussed in the following account.

1. Recordings—Flight events

The stimuli for the listening tests originated from 24 field recordings (i.e., 24 flight events) of takeoffs and landings of helicopters and propeller-driven aircraft on Grenchen Airport, Switzerland (ICAO code: LSZG).

Monophonic audio recordings were performed by means of a series of measurement microphones “B&K 4188” (Brüel & Kjær, Naerum, Denmark) and “Nor 1227” (Norsonic AS, Tranby, Norway) placed at 4 m height above the ground in the eastern side of the airport—from which the aircraft took off and landed—left and right of the extensions of the runway, as depicted in Fig. 3(b).

To auralize the flight events on 3D trajectories in the laboratory, positional data of the aircraft were recorded and time synchronized with the audio recordings. Position data were estimated with an optical tracking system consisting of two manually operated theodolites and triangulation. Figure 3 shows trajectories of the 24 flights chosen for this study in the horizontal plane and in the vertical direction (i.e., altitude). Only the relevant segments are shown which correspond to the duration of the final stimuli (24 s). Whereas landing profiles were much more similar (for these segments), takeoff profiles were scattered more. Specifically, helicopter takeoffs were at considerably higher altitudes.

Sound exposure levels $L_{AE,t10}$ (i.e., $L_{AE}$ calculated over the level measurement time “t10,” defined as the time interval in which the level $L_{AS}$ lies between $L_{AS,max}$ and $L_{AS,max} - 10$ dB) covered a range of 77–85 dB(A) for helicopters and of 64–85 dB(A) for propeller-driven aircraft. However, the $L_{AE,t10}$ range of propeller-driven aircraft landings were lower than propeller-driven aircraft takeoffs, helicopter takeoffs, and helicopter landings. To have comparable level ranges for all four combinations of source type and

![FIG. 2. (Color online) Hemispherical loudspeaker setup of the study.](image)

![FIG. 3. Trajectories for takeoffs [left: (a) and (c)] and landings [right: (b) and (d)]. Horizontal plane and altitude (Z) are shown above [(a) and (b)] and below [(c) and (d)], respectively. X, Y, and Z represent the distance from a reference point at the end of the runway, i.e., (0,0,0), while t_min indicates the time of the minimum distance between the recording microphones and the flying aircraft. In (b), microphone positions (original observer positions) are shown as asterisks and open circles represent virtually simulated observer positions (plus 300 m).](image)
procedure, different level adjustment techniques were applied: either a virtual microphone (i.e., observer) displacement of 300 m [see Fig. 3(b)] was carried out and/or simple level reduction (source level reduction) by 5 or 10 dB was used. This will be explained further in Sec. II C 2.

The 24 original recordings—12 helicopter events (6 takeoffs and 6 landings) and 12 propeller-driven aircraft events (6 takeoffs and 6 landings)—were selected carefully to exhibit as little external noise, unwanted dominant ambient sound, or microphone inherent noise, as possible.

2. Processing and auralization

Flight events underwent the following processing steps to prepare them for playback in AuraLab. Most of the processing—beside bass management—was carried out with MATLAB R2016b (MathWorks, Natick, MA).

- **Preprocessing**: in a first stage of processing, the 24 original recordings were calibrated, high-pass filtered ($f_c = 20$ Hz), and low-pass filtered ($f_c = 10$ kHz) to attenuate irrelevant noise (e.g., microphone inherent noise).

- **Optional propagation filtering**: in addition to the original measurement points, virtual observer positions were simulated for farther distances (plus 300 m), as depicted in Fig. 3(b). Modeling the aircraft as a moving point source, the simulation considered geometrical spreading, air absorption, and the Doppler frequency shift. Time-variant propagation filters were generated as a function of the source-observer-geometry and weather conditions (Heutschi et al., 2014; Pieren et al., 2016) and applied to the audio recordings. As a result of this, compared to the recordings, the resulting stimuli at the virtual observer positions had lower sound pressure levels, flatter $L_{AS}$ curves, weaker Doppler shift, lower elevation, lower azimuth, and greater attenuation of high frequencies components. Variations in ground effect, turbulence effects or radiation angle were not considered.

- **Optional level shift**: landings of propeller-driven aircraft exhibited generally lower levels than the three other combinations. To have a comparable level range, the stimuli of both versions—i.e., original and with propagation simulation—of three flight events (of other categories) were attenuated by 5 or 10 dB.

Furthermore, the two flight events with the lowest original levels—both landings of propeller-driven aircraft—did not undergo the propagation filtering. That is, 24 and 22 flight events were played back at a reference observer distance of approximately 150–200 and 450–500 m, respectively.

- **Spatialization**: to render the original monophonic recordings in 3D, the signals had to be allocated to the loudspeakers depending on the static position and direction of the observer, as well as on the time-variant position of the aircraft. Therefrom the time-variant angles of incident sound (azimuth and elevation) were computed. The 3D audio rendering algorithm is based on the common vector base amplitude panning (VBAP) method (Pulkki, 1997). By means of a crossover, the rendering algorithm works on two separate subbands. This allows distinctive normalizing factors for the panning. Subsequently, (panned) signal parts of the two subbands are mixed for each loudspeaker.

- **Adding background ambient sound**: Listening to the spatialized flight events in AuraLab revealed occasional microphone recording noise and/or dominant background ambient sounds (e.g., from singing birds) in a number of recordings. Since the auralization used a monophonic recording, such sounds would “fly with the auralized flight events from one loudspeaker to the next” on their flight trajectories. Therefore, vegetation background ambient sound was mixed with the flight events to cover the inherent background noise. Furthermore, as subjects were supposed to imagine that they were sitting in their garden, background nature/vegetation sounds helped to increase the degree of realism of the listening situation.

Background sounds originated from surround recordings done by means of an Ambisonic microphone Soundfield SP5200 (Soundfield, London, UK). A measurement microphone was used to calibrate the system and record the sound pressure level. Ambience recordings were decoded for a 2D (i.e., horizontal plane at subject’s ear level, i.e., $0^\circ$ of elevation) presentation in AuraLab. The individual loudspeaker signals were mixed with the corresponding loudspeaker signals of the auralized flight events. The original (and playback) level ($L_{Aeq}$) of the background ambient sounds was about 37.5 dB(A), which was considerably lower than the flight events ($L_{Aeq}$). They did not mask the flight events in any moment or frequency region.

3. Pilot listening test—Length of stimuli

The events of the processed stimuli covered event times ($t_{10}$) of 8 to 40 s (median 18 s, mean 19 s, and standard deviation 8 s). To avoid a possible bias of the ratings by varying stimuli length, a pilot experiment was carried out to determine an optimal fixed stimuli length for all flight events. The stimuli length was to be long enough to cover most of the relevant audible duration of all events, but still short enough to avoid fatigue of the subjects with an unnecessary long experiment. Given the two requirements to stimuli length, the pilot experiment was done in two parts, as specified below.

The test procedure and the software were similar to those of the main listening tests—which will be described later. Prior to the pilot experiment, participants were informed about the “noise annoyance” topic of the listening test, but not about the objective to determine an appropriate stimuli duration.

The test consisted of two parts. In part 1, a subset of the flight events with varying lengths was used to investigate whether stimuli length influenced the annoyance rating. Forty stimuli were used here, resulting from four flight events (P28A-takeoff, PA46-landing, A119-takeoff, and AS35-landing) at two observer distances (original and virtual) and with different lengths of 18, 24, 30, 36, and 42 s ($4 \times 2 \times 5 = 40$ Stimuli). For a given trial, subjects rated
their annoyance to the presented stimulus on the ICBEN\textsuperscript{2} 11-point scale (ISO/TS 15666, 2003); further details about the annoyance question and the rating scale will follow in Secs. II D and II E. In part 2, subjects’ preferences were investigated. After listening to each stimulus, subjects classified the stimulus length as to be “too short,” “spot-on,” or “too long,” for annoyance rating, which was coded as “–1,” “0,” “+1,” respectively, for the statistical analysis (Schäffer \textit{et al.}, 2016). Ten stimuli were used here, resulting from two flight events (P28A-takeoff at the original observer distance and AS35-landing at the virtual observer distance) presented in the five lengths mentioned above.

Twelve self-reported normal-hearing subjects (four females and eight males) participated in the pilot experiment. They were aged between 24 and 54 years (median 28 years). The pilot experiment took about 40 min for each subject.

The data of part 1 revealed no systematic effect of length (Fig. 4). There was no significant correlation between annoyance and stimulus length (Pearson’s $r = -0.06$, $p > 0.05$).

The data of part 2 showed a preference for a stimulus length of 24 s (Fig. 5). The effect of stimulus length on perceived length (i.e., preference) was significant ($p < 0.001$).

The results of the pilot listening tests revealed that, on the one hand, the stimuli length—in the range tested here—did not significantly affect annoyance ratings (part 1) and, on the other hand, the subjects found 24 s to be appropriate for the purpose of the study (part 2). Therefore, the length of stimuli for the main experiment was set to 24 s.

4. Final set of stimuli

The final set of 46 stimuli is listed in Table I. The stimuli covered four helicopter and five propeller-driven aircraft types, takeoffs and landings, as well as original and processed recordings as described above. As a measure of low-frequency content, mean $L_{CE} − L_{AE}$ is shown for each type-procedure category. Compared to helicopters, mean $L_{CE} − L_{AE}$ was slightly higher (around 1 dB) for propeller-driven aircraft. Furthermore, Fig. 6 shows $L_{AE}$ range [64–85 dB(A)] of the four group of stimuli as a function of their $L_{ASmax}$ [53–78 dB(A)].

More details on the stimuli can be found in Figs. 7 and 8, whereby mean sound exposure one-third octave spectra and mean $L_{AS}$ curves are shown for categories of source types and procedures. Figure 7 reveals tonal components for the noise of propeller-driven aircraft. In comparison, takeoffs of propeller-driven aircraft and helicopter landings exhibit the steepest and the flattest $L_{AS}$ curves, respectively (see Fig. 8). Furthermore, although an inspection of flight path velocities of the flight events (not shown here) revealed slightly steeper curves (i.e., higher acceleration) for helicopters than for propeller-driven aircraft, both for landings and for takeoffs, the ground speed ranges at the shortest distance of the different type-procedure categories were close to each other (i.e., roughly 27–52 m/s).

D. Annoyance ratings and questionnaire

Subjects rated their annoyance to the individual flight events on the ICBEN 11-point scale (ISO/TS 15666, 2003). Thereby, the following question was asked for every stimulus [in German, taken from Schäffer \textit{et al.} (2016)]: “When you imagine that this is the sound situation in your garden, what number from 0 to 10 represents best how much you would be bothered, disturbed or annoyed by it?”

A questionnaire complemented the listening tests. The first part contained questions about hearing (questions of the Swiss National Accident Insurance Fund, SUVA) and well-being. In the second part, questions on subjects’ attributes gender, age, living environment, and attitude were asked. Noise sensitivity was determined—besides by a direct question whether the subjects categorize themselves as “noise critical”—by means of the NoiSeQ-R [the reduced version of the “Noise-Sensitivity-Questionnaire” (Schutte \textit{et al.}, 2007)].

E. Test procedure

The experiments were conducted as focused listening tests, i.e., the subjects deliberately listened to the stimuli and rated them during or directly after playback. The stimuli were played once only, with a 1.2-s break between stimuli after complete playback.

Subjects did the tests individually. After reading information and signing a consent form to participate, they answered...
They were then introduced to the listening test and the software which guided them through the test and recorded their annoyance ratings. The setup is shown in Fig. 1. The subjects listened to three orienting and two training events before starting the main listening test.

The 46 flight events were presented in random order for each subject. The orientation of the virtual acoustical scene was balanced between the subjects; i.e., either takeoffs from right to left and landings from left to right, or vice versa. Furthermore, while for each subject, both background ambient sounds (BG 1 and BG 2) were used—each for half of the stimuli—the order of the background sounds was balanced between the subjects. That is, either BG 1 for odd and BG 2 for even playback orders, or vice versa.

After the listening test, the subjects filled out the rest of the questionnaire.

F. Subjects

Fifty-six subjects (22 females and 34 males) participated in the main experiment. All subjects declared to have normal hearing and to feel well. They were aged between 18 and 71 years (median 43 years). Forty subjects worked at a research institution in Dübendorf, Switzerland, either at the authors’ institution, Empa, or at the adjacent institution, Eawag. Furthermore, 39% of the subjects work in the field of acoustics/noise control.

The subjects covered a wide range of noise sensitivities with values of 0.8–2.9 (median 1.9). All 56 subjects could be included in the statistical analysis.

G. Statistical analysis

Statistical analysis was carried out with IBM SPSS Statistics, version 22 (IBM Corporation, Armonk, NY). Tested effects were considered significant if the probability, \( p \), of the observed results under the null hypothesis was <0.05.

1. Annoyance ratings

Associations of design parameters (aircraft type, procedure, and \( L_{AE} \)) with annoyance were analyzed by means of linear mixed-effects models (Pinheiro and Bates, 2000), combining fixed effects of categorical variables aircraft type and procedure, covariate \( L_{AE} \), random effects (subjects), and
their interactions to predict dependent variables annoyance ratings.

In addition to main design parameters and their interactions, the sequence of playback, subjects’ attributes, and different random effect models were studied regarding their possible significant link to annoyance. Thus, several models of different degrees of complexity were established. The models were compared using Akaike information criterion (AIC) (Akaike, 1973) and Bayesian information criterion (BIC) (Schwarz, 1978), where the model with the lowest AIC/BIC was preferred. Non-significant variables and interactions were excluded from the final model, which will be presented later in this paper.

Compliance with model assumptions was visually confirmed by means of residual plots. Goodness-of-fit of the final models were assessed with marginal ($R^2_m$) and conditional ($R^2_c$) coefficients of determination (Vonesh et al., 1996), representing the variance explained by the fixed factors and the variance explained by the fixed plus random factors, respectively. Quantification was done according to Nakagawa and Schielzeth (2013) and Johnson (2014).

2. Probability of HA

The dependent binary variable “highly annoyed” (“not HA” vs “HA”) was coded from annoyance ratings as follows (Miedema and Oudshoorn, 2001; Schultz, 1978): HA = 0, if annoyance rating ≤ 7; HA = 1, if annoyance rating ≥ 8.

The dependence of HA on the design parameters was analyzed by binary logistic regressions (Hosmer et al., 2013) which estimate the probability of HA ($p_{HA}$). We intended to establish exposure-response curves representing an average $p_{HA}$ within the population. Therefore, generalized estimating equations (Liang and Zeger, 1986) were used (srs procedure GENLIN) to account for the repeated ratings of the subjects, as they predict a population-averaged response (Hu et al., 1998; Schäffer et al., 2017). The same model parameters were included and tested here as for the annoyance.
propeller-driven aircraft seem negligible. This will be further analyzed later by means of the linear mixed-effects model in Sec. III B 1.

2. Rating times

Figure 9(b) indicates that rating time varied strongly between individual ratings and tended to be shorter at high $L_{AE}$. A weak but significant negative correlation between $L_{AE}$ and rating time confirms this ($r = -0.18$, $p < 0.001$). No systematic significant differences can be observed in Fig. 9(b) between rating times for helicopters and propeller-driven aircraft ($p > 0.05$), though the average rating time tended to be slightly higher for helicopters.

Furthermore, a weak negative correlation was observed between annoyance and rating time ($r = -0.251$, $p < 0.001$). The more annoying a stimulus was, the more quickly the annoyance was rated by the subjects.

3. Playback sequence

Figure 10 shows mean annoyance and mean rating time across subjects and pooled over source type and procedure as a function of playback sequence. On average, annoyance tended to increase with playback number. This will be discussed further in Sec. III B 1. Rating time correlated weakly and negatively with playback sequence ($r = -0.11$, $p < 0.001$). On average, subjects rated more quickly with increasing playback sequence. On the one hand, learning effects might have helped subjects to rate more quickly (i.e., more confidently) as they listened to more excerpts. On the other hand, subjects might have rated faster because they found the test tiresome and they wanted to finish it more quickly.

B. Evaluation of the effects on annoyance

Mean short-term annoyance ratings across subjects are shown in Fig. 11 as a function of sound exposure level ($L_{AE}$) for the individual stimuli. For the larger part of $L_{AE}$, no significant difference was found between annoyance ratings for two aircraft types at comparable $L_{AE}$. However, whereas, at low $L_{AE}$, propeller-driven aircraft were rated as slightly more annoying than helicopters, at high $L_{AE}$ the opposite was observed (i.e., helicopters were slightly more annoying than propeller-driven aircraft). In both cases, the differences were less than one annoyance unit (Fig. 11).

1. Linear mixed-effects model of annoyance

The following linear mixed-effects model was found to be optimal to predict annoyance:

$$y_{ijk} = \mu + \tau_{Src,i} + \tau_{Pro,j} + \beta \cdot L_{AE,ij} + \gamma \cdot S_{ijk} + \tau_{Src \times Pro,ij} + \beta_{Src,i} \cdot L_{AE,ij} + u_{0k} + u_{1k} \cdot L_{AE,ij} + \epsilon_{ijk}. \quad (1)$$

In Eq. (1), $y_{ijk}$ is the dependent variable annoyance, $\mu$ is the overall grand mean, $\tau_{Src,i}$ and $\tau_{Pro,j}$ denote the categorical variables source type (two levels: $i = 1, 2$) and procedure (two levels: $j = 1, 2$), $L_{AE,ij}$ and $S_{ijk}$ are the continuous variables $L_{AE}$ and playback sequence, and $\beta$ and $\gamma$ are regression coefficients. Further, $\tau_{Src \times Pro,ij}$ and $\beta_{Src,i}$ represent interactions of source type with procedure and $L_{AE,ij}$, respectively. The (unstructured) random effect terms $u_{0k}$ and $u_{1k}$ are participants’ random intercept and slope—depending on $L_{AE}$—for $k = 1, \ldots, 56$. Finally, the error term $\epsilon_{ijk}$ is the random deviation between observed and expected values of $y_{ijk}$. All parameters contributed significantly to the model (all $p < 0.05$). Model coefficients are shown in Table II. The model of Eq. (1) explains a large part of the variance of the annoyance data ($R^2 = 0.80$). The fixed effects alone explain almost half of total variance in the data ($R^2_m = 0.46$).

Equation (1) indicates a significant interaction between source type and procedure in predicting annoyance, whereby helicopter takeoffs tended to be less annoying than helicopter landings, while no significant difference was found between annoyance reactions to takeoff and landing of propeller-driven aircraft (Fig. 12).

2. Runway direction and order of background sounds

It was mentioned in Sec. II E that runway direction and order of background sounds were balanced between subjects. None of the two variables had a significant contribution in the linear mixed-effect model (all $p > 0.05$). They did not affect annoyance significantly.

3. Personal attributes

It was investigated whether including personal attributes would improve the model. Subjects’ age and gender did not significantly influenced annoyance (all $p > 0.05$). On average, annoyance rating of subjects who worked in the field of acoustics/noise control was higher than the rest of the subjects, however, only slightly (i.e., 0.5 unit on the 11-point annoyance scale). Consequently, including this parameter in the model showed no significant effect ($p = 0.19$). Therefore, similar results would be expected with a random panel.

Ratings of the subjects who lived near an airport or airfield were, on average, lower than those of other subjects.
However, included in the model, this variable did not contribute significantly \( (p > 0.05) \) to explaining the observed variance in the data.

Subjects, who stated that the Swiss politics should pay more attention to quietness and protection of environment than to the Swiss economy, were, on average, more annoyed by flight events in the laboratory than subjects who stated the opposite. However, including this variable did not improve the model significantly \( (p > 0.05) \).

Subjects were asked which aircraft category (helicopters or propeller-driven aircraft) they experienced in their everyday life “more frequently” and which aircraft category they generally found “louder” or “more annoying.” 68%, 66%, and 68% of the subjects stated to experience helicopters more frequently, find them louder and more annoying, respectively. It was investigated whether their statements related to their annoyance ratings for these source types in the laboratory. No relationship was found between the annoyance ratings and any of the three variables (all \( p > 0.05 \)).

Noise sensitivity measured by NoisSEQ-R did not significantly contributed to the model \( (p > 0.05) \). However, subjects, who stated (directly) to be “noise critical,” were, on average, more annoyed by the flight events in the laboratory. Including this variable improved the model significantly \( (p < 0.05) \). Doing so, although the fixed effects of the model explained 3\% more variance in the data \( (R^2_m = 0.49) \), in total, the model did not explain more variance \( (R^2_m = 0.80) \) than when this variable was not included; see above for comparison.

### C. Binary logistic regression model for the probability of HA (pHA)

The following binary logistic regression model was established to predict the observed relative frequency of HA:

\[
\text{logit}(pHA) = \mu + \tau_{\text{Src},i} + \tau_{\text{Pro},j} + \beta \cdot L_{AE,ij} + \tau_{\text{Src} \times \text{Pro},ij} + \beta_{\text{Src},i} \cdot L_{AE,ij}.
\]  

(2)

In Eq. (2), \( \text{logit}(pHA) = \ln(pHA/[1 – pHA]) \) is the logit for \( pHA \) \citep{Taghipour2019}. All other variables have the same notation as in Eq. (1). In contrast to the linear mixed-effects model [Eq. (1)], the effect of the playback sequence was not significant (hence, not included). The model of Eq. (2) predicts the individual ratings satisfyingly, with a coefficient of discrimination of 0.34 \citep{Tjur2009} and rate of

### TABLE II. Model coefficients (Coeff.), with 95\% CI and probabilities \( (p) \) of the linear mixed-effects model for annoyance and of the population-averaged logistic regression model for probability of HA, and odds ratio \( (OR = \exp(\text{Coeff.}) \) with 95\% CI for the logistic regression model. The parameters and symbols are explained in Eqs. (1) and (2).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Linear mixed-effects model, Eq. (1)</th>
<th>Coeff.</th>
<th>95% CI</th>
<th>OR</th>
<th>Population-averaged logistic regression model, Eq. (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>( \mu )</td>
<td>( -15.2987 )</td>
<td>(-17.5364; -13.0610)</td>
<td>0.00</td>
<td>(-21.2890 )</td>
<td>(-25.6754; -16.9026)</td>
</tr>
<tr>
<td>Source type*</td>
<td>( \tau_{\text{Src},j} = \text{heli} )</td>
<td>(-1.8113 )</td>
<td>(-3.0997; -0.5228)</td>
<td>0.01</td>
<td>(-3.9167 )</td>
<td>(-6.9568; -0.8765)</td>
</tr>
<tr>
<td>( \tau_{\text{Src},j} = \text{prop} )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
</tr>
<tr>
<td>Procedure*</td>
<td>( \tau_{\text{Pro},j} = \text{lan.} )</td>
<td>(-0.0805 )</td>
<td>(-0.2217; 0.0607)</td>
<td>0.26</td>
<td>(-0.0301 )</td>
<td>(-0.2683; 0.2080)</td>
</tr>
<tr>
<td>( \tau_{\text{Pro},j} = \text{t.o.} )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
</tr>
<tr>
<td>( L_{AE} )</td>
<td>( \beta )</td>
<td>0.2845</td>
<td>[0.2557; 0.3132]</td>
<td>0.00</td>
<td>0.2719</td>
<td>[0.2174; 0.3265]</td>
</tr>
<tr>
<td>Playback no.</td>
<td>( \gamma )</td>
<td>0.0057</td>
<td>[0.0024; 0.0090]</td>
<td>0.00</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Source ( \times ) Procedure.</td>
<td>( \tau_{\text{Src} \times \text{Pro},ij} = \text{heli \times lan.} )</td>
<td>0.4579</td>
<td>[0.2717; 0.6440]</td>
<td>0.00</td>
<td>0.4490</td>
<td>[0.1405; 0.7574]</td>
</tr>
<tr>
<td>( \tau_{\text{Src} \times \text{Pro},ij} = \text{heli \times t.o.} )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
</tr>
<tr>
<td>( \tau_{\text{Src} \times \text{Pro},ij} = \text{prop \times lan.} )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
</tr>
<tr>
<td>( \tau_{\text{Src} \times \text{Pro},ij} = \text{prop \times t.o.} )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
</tr>
<tr>
<td>Source ( \times L_{AE} )</td>
<td>( \beta_{\text{Src},i} = \text{heli} )</td>
<td>0.0209</td>
<td>[0.0041; 0.0377]</td>
<td>0.01</td>
<td>0.0473</td>
<td>[0.0081; 0.0864]</td>
</tr>
<tr>
<td>( \beta_{\text{Src},i} = \text{prop} )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
<td>0 ( ^* )</td>
</tr>
</tbody>
</table>

*a heli = helicopter; prop = propeller.

*b Redundant coefficients are set to zero.

*c lan. = landing; t.o. = takeoff.
correct HA predictions of 79%. Model parameters are listed in Table II.

Figure 13 shows relative frequency of HA and the model prediction for \( p_{HA} \) as a function of \( L_{AE} \). A comparison of \( p_{HA} \) shows differences between helicopter and propeller-driven aircraft only at low and high \( L_{AE} \), as is the case for annoyance (see Fig. 11). The effect of the interaction between source type and procedure was similar as for annoyance (see Fig. 12).

IV. DISCUSSION

For the major range of \( L_{AE} \), no significant aircraft type dependent differences were observed between short-term annoyance ratings of helicopters and propeller-driven aircraft. However, whereas at low \( L_{AE} \), the propeller-driven aircraft were slightly more annoying than helicopters, at high \( L_{AE} \), the opposite was observed. It has to be noted that no helicopter event of levels below 66 dB(A) and only one propeller-driven aircraft event (takeoff) of levels above 83 dB(A) were tested in this study (see Table I and Fig. 6). Hence, the predictions of the model at the lowest and highest levels should be interpreted with caution. The observed level differences at equal annoyance ratings (i.e., possible level correction for or against helicopters) were below 1 dB. The corresponding differences in annoyance ratings observed in this study were below 0.5 points, which is less than the resolution of the 11-point scale (i.e., 1.0 point).

No significant differences were observed between annoyance ratings (and also \( p_{HA} \)) for takeoffs and landings of propeller-driven aircraft at equal \( L_{AE} \). For helicopters, however, landings were, on average, more annoying than takeoffs. This is in accord with the observations by Schomer (1983). For the stimuli in the experiment presented here, one-third octave spectra (see Fig. 7) do not indicate clear

energy differences between helicopter takeoffs and landings. However, altitude curves in Fig. 3 show steeper slopes for helicopter takeoffs than for helicopter landings. On average, helicopter landings exhibit flatter \( L_{AS} \) curves and therefore longer t10 than helicopter takeoffs. That is, while all stimuli had a length of 24 s, on average, subjects listened to longer portions of sounds close to \( L_{ASmax} \) when helicopters were landing than when they were taking off.

Since literature mostly compared helicopters to jet aircraft (and not to propeller-driven aircraft), comparability of the present results is difficult. However, literature either suggests that helicopters are more annoying than other types of aircraft, or that there is no difference. The present study, thus, seems in accordance with studies which suggest to assess helicopters noise as those of fixed-wing aircraft (Atkins et al., 1983; Berry et al., 1979; Fields and Powell, 1987; Gjestland, 1994; Janssen et al., 2017; Mestre et al., 2017; Molino, 1982; Newman and Beattie, 1985).

In addition to analyzing the results with respect to \( L_{AE} \), a similar approach was applied considering \( L_{ASmax} \) or \( L_{CE} \). Since \( L_{AE} \), \( L_{ASmax} \), and \( L_{CE} \) were highly correlated (all \( r > 0.9 \), all \( p < 0.001 \)), similar patterns were found. The level was the major predictor, no matter whether expressed in \( L_{AE} \), in \( L_{ASmax} \), or in \( L_{CE} \). However, the correlation between short-term annoyance ratings and \( L_{AE} \) (\( r = 0.67, p < 0.001 \)) was slightly stronger than with \( L_{ASmax} \) (\( r = 0.65, p < 0.001 \)) or with \( L_{CE} \) (\( r = 0.62, p < 0.001 \)). Linear mixed-effects models, in which \( L_{ASmax} \) or \( L_{CE} \) were included instead of \( L_{AE} \), showed higher BIC (Schwarz, 1978) and AIC (Akaike, 1973) values, which concludes that \( L_{AE} \) was a better predictor of short-term annoyance than \( L_{ASmax} \) or than \( L_{CE} \). This might be different for heavy-weight helicopters or aircraft.

Questioning subjects’ everyday life experience with helicopters and propeller-driven aircraft revealed that helicopters were experienced more often and were found louder and more annoying than propeller-driven aircraft. This,

![FIG. 12. Mean short-term annoyance across subjects as a function of the \( L_{AE} \) for the four source-procedure combinations. The lines show the estimates of the linear mixed-effects model for each combination.](image)

![FIG. 13. Exposure-response curves: relative frequency and \( p_{HA} \) as a function of \( L_{AE} \) for helicopter (filled lines and circles) and propeller-driven aircraft (dashed lines and triangles).](image)
however, was not found to be related to subjects’ ratings in the laboratory. It might be, that, in their everyday life, subjects experience helicopter noise from shorter propagation distances, and therefore louder/more annoying. However, this would be covered by $L_{AE}$ (or $L_{dn}$) in the process of noise assessment. That is, the question is whether they find helicopters more annoying, but, whether annoyance differences can be explained by $L_{AE}$, sufficiently. This was the case for the presented laboratory experiment, for the major part of $L_{AE}$ range.

Fidell et al. (2011) introduced the so-called community tolerance level, CTL, which is calculated based on $L_{dn}$ and shows the sound pressure level at which 50% of the population is highly annoyed. For the data presented here, an averaged sound pressure measure (such as $L_{dn}$) would not be applicable. Nevertheless, Fig. 13 shows that an $L_{AE}$-based CTL was equal to 78.5 dB(A). As the single flight events here and their exposure levels are not comparable with any averaged data reported in other studies cited in this paper, a comparison is not possible. The two laboratory experiments reported here which used individual flight events (Gjestland, 1994; Janssen et al., 2017) were carried out by means of paired comparisons, whereby no ICBEN scales were used and no pHA was reported.

Since short-term annoyance was investigated in the present study by means of a focused laboratory experiment, it is important to note the differences to the annoyance observed in the field. Long-term annoyance is not only influenced by the source and its sound pressure level, but also by psychological, political, cultural, and social aspects, such as political/business announcements (Guski and Bosshardt, 1992; Schreckenberg et al., 2017). For example, when it is announced that an airport will be expanded by a new runway, or alternatively more flight routes/destinations in future, annoyance reaction to flight events might increase, even before any additional flight is added (Schreckenberg et al., 2017). It was mentioned before that, generally, one should distinguish between “short-term annoyance” (Bolin et al., 2014) or “psychoacoustic annoyance” (Fastl and Zwicker, 2007) and long-term annoyance assessed in field surveys (Guski and Bosshardt, 1992). On the other hand, one of the advantages of a laboratory setup is the degree of control in the experimental design and execution, as a result of which all subjects experience the same scenario and equipment. This is very helpful in comparison studies, such as the one presented here.

Further, to enhance the ecological validity and representativeness of the present study, an effort was made to recruit participants who live near an airfield. 66% of the subjects lived close to an airfield, the vast majority of whom lived nearby Zurich International Airport (ICAO code: LSZH) or Dübendorf Air Base (ICAO code: LSMD), both in Switzerland.

Long-term—as well as short-term—annoyance to helicopters might also be affected by building vibrations. Schomer and Neathammer (1987) and Schomer and Averbuch (1989) investigated the role of building vibrations for the perceived annoyance to helicopters inside buildings. This raised an open question to whether only the $L_{dn}$ is responsible for the perceived annoyance or also hearing windows, ceiling tiles, or objects in the room rattle and/or the general perception of building vibration (Schomer, 1983; Schomer and Averbuch, 1989; Schomer and Neathammer, 1987). In a recent laboratory experiment Janssen et al. (2017) investigated exposure-response relationships for helicopters compared to those for jet fighter aircraft. In particular, the study focused on the role of helicopter noise-induced rattle in increasing the annoyance response. Depending on the dwelling and the type of helicopters, less or more indoor vibrations and rattle were observed caused by the low frequency noise of the helicopters. The jet aircraft, however, did not cause such vibrations. It was found that the measured differences in outdoor and indoor levels were not similar for types of aircraft and dwellings. At similar A-weighted indoor levels no significant difference was found between the annoyance of helicopters and jet fighter aircraft, which is in accord with the results of the current study. Further, because the rattles were caused by the low frequency noise of helicopters and there were considerable differences between outdoor and indoor levels depending on the isolation in the dwelling and type of helicopter, Janssen et al. (2017) argued that C-weighted noise levels might be more suitable to take both of these aspects (rattle and annoyance) into account, because they better reflect the low frequency components. However, similar to the finding of the present paper, they also mentioned that C-weighted levels proved less able than A-weighted levels to predict annoyance differences between helicopter and jet fighter aircraft noise (Janssen et al., 2017). It should be mentioned that helicopters of the experiment by Janssen et al. (2017) were considerably heavier than those of the present study (civil helicopters of below 3200 kg; see Table 1), which might have led to higher C-weighted levels. In general, heavy military transport helicopters might exhibit higher low frequency energy than the ones used in the present study. Investigations in the role of low-frequency content of the noise and the building vibrations on annoyance were/are beyond the scope of this paper and will not be addressed here further.

V. CONCLUSIONS

In this laboratory experiment, focused listening tests were carried out to investigate short-term annoyance to noise of civil helicopters and propeller-driven aircraft. Takeoffs and landings of these two aircraft classes were presented for a broad range of sound exposure levels and two observer distances. The sound exposure level ($L_{AE}$) was found to be the main predictor of noise annoyance. Source type was not a predictor of noise annoyance inherently when no other interactions were considered. However, there was a significant interaction of source type and procedure, whereby helicopter landings were slightly more annoying than helicopter takeoffs. Because the observed differences were small, the authors conclude that the results do not suggest a need for a source dependent assessment of noise annoyance between these two source types.
ACKNOWLEDGMENTS

This study was funded by the Swiss Federal Office for the Environment (Assignment No. 5211.01228.100.01). The authors are very grateful to the participants of the preliminary and main experiments of this study. They very much appreciate the cooperation with the “Institute of Geodesy and Photogrammetry” of the ETH Zürich, who recorded the synchronized position data of the flight events, and would like to thank Felix Schlatter for measurement data and analysis of the original recordings. Further, they are thankful to Mark Brink, Kornel Köstli, and Jean Marc Wunderli for their valuable inputs.

1The background was a 5-dB correction (penalty) for helicopter noise assessment according to the Norwegian regulations at the time (Gjestland, 1994).
2International Commission on Biological Effects of Noise.
3Swiss Federal Institute of Technology in Zurich.


