Evaluation of flyover auralizations of today's and future long-range aircraft concepts

Beat Schäffer[1]
Empa, Swiss Federal Laboratories for Materials Science and Technology
Überlandstrasse 129, 8600 Dübendorf, Switzerland

Lothar Bertsch[2]
German Aerospace Center (DLR)
Bunsenstraße 10, 37073 Göttingen, Germany

Ingrid Le Griffon[3]
Office National d'Études et de Recherches Aérospatiales (ONERA)—Paris Saclay University
29, avenue de la Division Leclerc – BP 72 – 92322 Châtillon Cedex, France

Axel Heusser[4]
Empa, Swiss Federal Laboratories for Materials Science and Technology
Überlandstrasse 129, 8600 Dübendorf, Switzerland

Catherine Lavandier[5]
ETIS laboratory, CY Cergy Paris University, ENSEA, CNRS, UMR8051
95000 Cergy, France

Reto Pieren[6]
Empa, Swiss Federal Laboratories for Materials Science and Technology
Überlandstrasse 129, 8600 Dübendorf, Switzerland

ABSTRACT

The European research project ARTEM (Aircraft noise Reduction Technologies and related Environmental iMpact) develops innovative aircraft noise reduction technologies such as advanced engine fan acoustic lining, metamaterials and low-noise high-lift systems applied to a vehicle with enhanced capabilities for shielding of the engine noise, namely, a blended wing body. Using aircraft flyover auralization in laboratory listening experiments, such future technologies can be evaluated with respect to human sound perception. To assess the reliability of such perception-based

[1] beat.schaffer@empa.ch

[2] lothar.bertsch@dlr.de

[3] ingrid.legriffon@onera.fr

[4] axel.heusser@empa.ch

[5] catherine.lavandier@cyu.fr

[6] reto.pieren@empa.ch

evaluations, the simulation chain should be validated with existing aircraft flyovers. This contribution presents a systematic and rigorous hierarchical validation of auralizations of current (today's) jet aircraft by means of comparisons with field recordings. Uncertainty in the source modelling is considered by using two different prediction tools for partial sound sources. In addition to comparing computed noise indicators, a psychoacoustic validation is done in laboratory listening experiments with a 3D loudspeaker array. The validation comprises three levels: (i) direct comparison of auralizations with recordings to study the identifiability of auralizations, (ii) ranking of auralizations and recordings regarding plausibility, and (iii) subjective annoyance ratings to test whether auralizations and recordings differ with respect to noise effects. Further, first results on the comparison of a future concept with a current aircraft are presented.

1. INTRODUCTION

Aircraft noise affects millions of people worldwide. In the year 2017, for example, 4 million people in Europe were estimated to be exposed to aircraft noise with a day–evening–night level (L_den) of 55 dB or higher [1]. With growing population and number of (post-COVID) aircraft movements likely to increase in the future (see https://www.eurocontrol.int/covid19; accessed 04/28/2022), this problem is likely to become even more pronounced in future.

To provide mitigation measures, the International Civil Aviation Organization (ICAO) introduced the "Balanced Approach" [2], which addresses aircraft noise management problems with four principal elements: noise reduction at the source, land-use planning and management, noise abatement operational procedures, and operating restrictions. Noise reduction at the source is a particularly effective measure. Indeed, massive improvements were obtained in the past (e.g., [3]). Future low-noise aircraft designs might thus further mitigate the aircraft noise problem.

The noise situation is usually assessed with conventional noise metrics such as the A-weighted equivalent continuous sound pressure level (L_Aeq), the L_den, or the Effective Perceived Noise Level (EPNL) used for noise certification purposes [4]. However, to obtain a more holistic picture, evaluating the situation also on the basis of human sound perception or noise effects is desirable [5], particularly already in the design phase of future aircraft technologies. Here, auralization (see, e.g., [6]) comes into play. Analogous to visualization, it allows creating virtual realities to listen to sound situations that do not necessarily exist in reality yet. Combined with listening experiments, this allows for a perception-based evaluation of noise reduction measures to complement conventional noise metrics. A recent study proved the feasibility of this approach for future low-noise aircraft technologies [7]. This approach is further developed and applied in the current work.

The European research project ARTEM (Aircraft noise Reduction Technologies and related Environmental iMpact) develops innovative aircraft noise reduction technologies such as advanced engine fan acoustic lining, metamaterials and low-noise high-lift systems applied to a vehicle with enhanced shielding of the engine noise, namely, a blended wing body. Using aircraft flyover auralization in laboratory listening experiments, such future technologies are evaluated with respect to noise annoyance. To assess the reliability of such perception-based evaluation, the simulation chain is validated in a first step with existing aircraft flyovers.

This contribution presents a systematic and rigorous hierarchical validation of auralizations of current (today's) jet aircraft by means of comparisons with field recordings. Further, first results on the comparison of a future aircraft concept with a current aircraft are presented.

2. METHODOLOGY

2.1 Concept

Within ARTEM, flyovers of future aircraft are auralized and used in laboratory listening experiments to evaluate their noise impact on humans. Two simulations based on different parametric prediction tools for partial sound sources, namely CARMEN from ONERA [8] and PANAM from DLR [9], are used to consider the uncertainty in the source modelling. In addition to comparing computed noise indicators (e.g., the A-weighted sound exposure level L_AE or psychoacoustic loudness

[10]), a psychoacoustic validation is done in dedicated laboratory listening experiments with a 3D loudspeaker array to assess the quality of the auralizations of existing aircraft regarding the identifiability of synthesized sounds, their plausibility, and their effects on humans (here, annoyance). The chosen approach thus validates the whole simulation chain (prediction tools–auralization), but not the simulations by the prediction tools and the auralization separately. However, the latter is not the objective of the validation, as the simulation chain needs to be valid as a whole. Also, both prediction tools were assessed in a large benchmark test [11] and compared with experimental data from flyover measurements (e.g., [12]), and Empa proved in various past auralization projects that their auralizations sound plausible (e.g., for wind turbines [13]). Note that the auralizations' whole simulation chain was fully blind to the field measurements, computed with independent calculations models and without any tuning to the measurements.

As design variables for the experiments, two existing aircraft types (Airbus A320-214 and A430-313), two procedures (departure, approach) and three origins of stimuli (reference, i.e., measurements [denoted as Ref. in the following]; simulation 1 [Sim. 1] and simulation 2 [Sim. 2]) were used. This resulted in 12 ($= 2 \times 2 \times 3$) acoustical stimuli of flyovers (4 recordings and 8 syntheses). The syntheses were enriched with recorded birdsong to adjust to the ambient situation of the recordings. The recordings were taken from a previous field measurement campaign around Zurich airport, Switzerland, from microphones located close to the noise certification points [14]. From the campaign, also meteorological and flight data recorder (FDR) data was available, yielding the necessary input data (e.g., flaps setting, low compressor speed N1, flight trajectory etc.) for the simulations and thus auralizations of the measured reference flights.

In the following, the methodology for the psychoacoustic validation is presented. The noise emission prediction of future aircraft technologies with the parametric system noise prediction tools are presented in a companion paper to this study [15]. In the following, we describe how virtual flyovers are auralized for receiver positions near the ground based on this input (Section 2.2) and how the flyovers are validated in laboratory listening experiments (Section 2.3).

2.2. Auralization

The auralization procedure is divided into three consecutive steps: (i) synthesizing time histories of emitted sound pressure at instantaneous emission angle, (ii) propagation simulation to observer location, and (iii) 3D spatial rendering and reproduction for spatial impression of virtual flyover.

In the first step, the emitted noise of each of the aircraft's sub-sources is synthesized as a function of the dynamic emission data and radiation angle. Here, the sound of broadband, non-harmonic sources (i.e., airframe and jet noise) and tonal components (engine fan and buzz saw noise) are generated. All sub-source signals are summed up to obtain a single source signal of the sound emission of a particular moving aircraft in the direction of a selected static observer point. In the second step, the effects of sound propagation from source to virtual observation points are simulated, namely, Doppler frequency shift, geometrical spreading, ground reflection, air absorption, and atmospheric turbulence-induced amplitude fluctuations and coherence loss. As a last step, the sound pressure signal is prepared with an amplitude panning technique to be fed to a calibrated hemispherical loudspeaker array in Empa's listening test facility AuraLab. This allows for creating a spatial impression of a flyover under laboratory conditions.

Sound synthesis (steps 1 and 2) was done with Empa's tool AURAFONE, which is described in [7]. Within ARTEM, AURAFONE was improved to account for atmospheric turbulence in ground effect [16]. Details on the 3D audio rendering and on AuraLab can be found in [17].

2.3. Listening experiments

In listening experiments, a systematic and rigorous hierarchical validation of auralizations of current jet aircraft using field recordings is done. The validation comprises three levels (Parts I–III):

- **Part I – Direct comparison of synthesized with recorded stimuli**: The syntheses and recordings are directly compared to study the identifiability of auralizations. To that aim, the participants do

pairwise comparisons (Ref. vs. Sim. 1, Ref. vs. Sim. 2; two-alternative forced choices, 2-AFC) to identify the recording, for each of the four presented situations (2 aircraft types × 2 procedures; cf. Section 2.1), in total eight comparisons. The data is analysed with a χ^2 test to assess whether the stimuli are classified into mutually exclusive classes ("Synthesis" and "Recording") or not.

- **Part II – Ranking of auralizations and recordings regarding plausibility of the stimuli**: The syntheses and recordings are ranked by the participants for groups of three stimuli (Ref., Sim. 1 and Sim. 2; three-alternative forced choice, 3-AFC) regarding plausibility, for each of the four presented situations, resulting in a total of four rankings. The data is analyzed again with a χ^2 test to assess whether syntheses and recordings are perceived as equally or differently plausible.

- **Part III – Subjective annoyance ratings to test whether auralizations and recordings differ with respect to noise effects**: The syntheses and recordings are rated with subjective noise annoyance using the ICBEN 11-point scale of ISO/TS 15666 [18] (direct scaling). The data is analyzed by mixed-effects models to test whether the syntheses and auralizations are equally or disparately annoying.

Laboratory procedure: The listening tests are conducted in single, individual sessions as focused experiments, i.e., with the participant deliberately listening to the stimuli. The test procedure follows the procedure described in [19]. In short, it consists of (i) an introduction to the research topic, (ii) filling out a consent form for study participation, (iii) a questionnaire about self-reported hearing capability and well-being as inclusion/exclusion criteria for study participation, (iv) the actual listening experiments (in three counterbalanced parts), with an instruction, orientation (example stimuli, only Part III), exercise ratings and main experiment, and (v) a post-experimental questionnaire with questions on subjects' characteristics such as sex, age or noise sensitivity (questionnaire NoiSeQ-R [20]). The latter covers values of 0 (noise-insensitive) to 3 (highly noise-sensitive). The experiments were approved by the ethics committee of Empa.

As the resulting acoustical stimuli of flyovers used in this study are quite loud ($L_{AF,max}$ of up to ~92 dB), a general level reduction by 5 dB was applied to all 12 original stimuli to prevent possible hearing damage. Further, in experimental Part III, the same stimuli were played back a second time with an additional attenuation of 15 dB (total attenuation of 20 dB, denoted as "attenuated stimuli") to prevent a ceiling effect in the annoyance ratings.

Participants: 31 participants (12 females, 19 males), aged 20–61 years (median of 36 years) with a noise sensitivity of 1.1–2.9 (median of 2.0) participated in the experiment.

3. RESULTS AND DISCUSSION

3.1 Computed noise indicators

Table 1 presents the L_{AE} of the 12 stimuli. Depending on the situation, the L_{AE} varies by ~10 dB, from ~85–95 dB. Auralized data agree very well with recordings, with a mean difference between syntheses and recordings of –0.9 dB (standard deviation ±1.5 dB). Similar agreements were also assessed for other noise metrics such as the EPNL. Also the psychoacoustic loudness according to Zwicker [10] agrees well, with the 5% percentile value of the syntheses deviating on average by –10% from the recordings. These relatively small differences are a first indication that the auralizations correctly reproduce the loudness sensation.

Also time histories (L_{AF} and loudness) and spectrograms of syntheses and measurements agree well (not shown). In fact, the syntheses cover the relevant frequency range for aircraft noise of 30–10'000 Hz, as well as all sound emission and propagation phenomena that are also caught by the measurements, e.g., spectral components due to the jet engines' fan tones, buzz saw noise components for departures, pitching down of tones due to the Doppler effect, or the time-varying comb filtering effect resulting from the interference of the direct sound with the reflection from the ground. Due to the parametric nature of the system noise prediction tools, operational changes along the flights are accounted for and can be identified in the results, e.g., configurational changes along approach flights.

Table 1: Sound exposure levels L_{AE} in dB of all stimuli of the validation study.

Situation	A320-214 Departure	A320-214 Approach	A340-313 Departure	A340-313 Approach
Measurement (Ref.)	84.8	92.8	94.2	94.5
Simulation 1 (Sim. 1)	85.4	90.5	92.1	94.8
Simulation 2 (Sim. 2)	86.4	91.1	91.8	93.3

The high agreement is surprising since the auralizations were produced fully blind to the acoustical field measurements, computed with independent calculations models and without any tuning to the measurements. Nevertheless, the spectrograms indicate that some spectral and temporal audible differences still exist (not shown). This is explored in the next section.

3.2 Psychoacoustic validation

Figures 1–3 below show the results pooled over the four situations (2 aircraft types × 2 procedures; cf. Section 2.1).

Figure 1 shows the relative frequencies of the stimuli classified by the participants as recording or synthesis in experimental Part I. Ideally, i.e., if measurements and simulations were not discriminable from each other, both classes ("Recording", "Synthesis") would have the same relative frequency of 50% for Ref. and Sim. 1 or Sim. 2. This would indicate that in 50% of the comparisons, Sim. 1 and Sim. 2 were (falsely) rated as a recording and Rec. as synthesis. Here, Sim. 1 was rated as recording in ~40% of the comparisons and Sim. 2 in ~20% ("false positive" according to detection theory). While the performance was thus quite good, particularly for the stimuli of Sim. 1, overall (pooled over design variables and comparisons) the differences in the relative frequencies of the two classes are significant, for both Sim. 1 vs. Ref. and Sim. 2 vs. Ref. (χ^2-tests, $p < 0.001$).
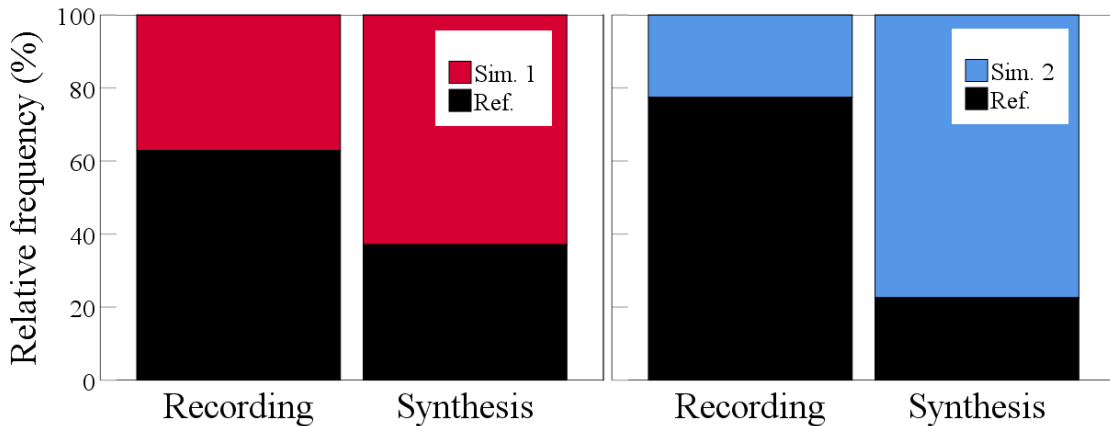

Figure 1: Experimental Part I: Relative frequencies of the reference (Ref.) and Simulation 1 (Sim. 1) (left), and Ref. and Sim. 2 (right) classified by the participants as "Recording" or "Synthesis".

Figure 2 shows the relative frequencies of the participants' ranking of simulations and recordings regarding plausibility in experimental Part II. Analogous to Part I, the three classes ("most plausible", "middle", "least plausible") ideally would have the same relative frequency of ~33% for Ref., Sim. 1 and Sim. 2. Here, in line with Part I, Ref. (i.e., recording) was rated as most and Sim. 2 as least plausible, but also Sim. 1 and Sim. 2 were partly rated as most plausible. Nevertheless, the differences in the relative frequencies of the three classes are overall significant (χ^2-test, $p < 0.001$).
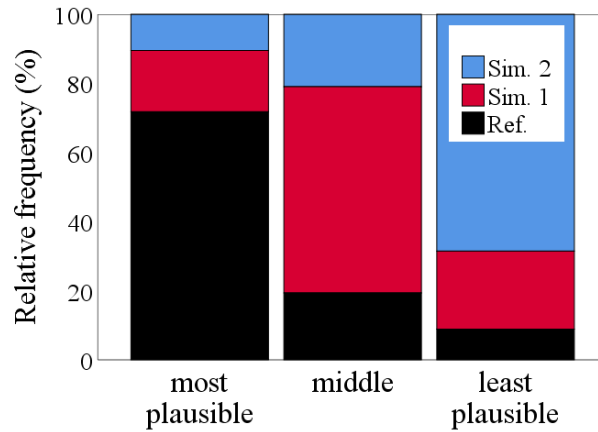
Figure 2: Experimental Part II: Relative frequencies of the reference (Ref.) and Simulations 1 and 2 (Sim. 1 and Sim. 2) of the plausibility ranking by the participants.

Finally, Figure 3 shows box-and-whisker plots of the individual annoyance ratings of experimental Part III. The ratings are very similar, independently of whether the stimuli originated from recordings (Ref.) or simulation (Sim. 1 and Sim. 2), and overall, both simulations performed similarly well. In fact, mixed-effects modelling analysis revealed no significant differences between the simulations and Ref. for the attenuated stimuli ($p = 0.57$, Figure 3 right). For the original stimuli (Figure 3 left), a significant difference was found ($p = 0.02$); however, a post hoc analysis with Bonferonni correction revealed that only Sim. 1 and Sim. 2 differed from each other ($p < 0.02$), but not from Ref. ($p > 0.39$). Figure 3 further reveals relative large scattering of the annoyance ratings. Part of the scatter is explained by the ~10 dB variation in the L_{AE} (cf. Table 1), and to some lesser extent by noise sensitivity, with noise-sensitive persons tending to be more annoyed. Besides, it reflects individual differences in the ratings, which was also observed in other studies and may be accounted for in mixed-effects modeling analysis (e.g., [19]). Further, one participant rated several stimuli with an annoyance of "0" (cf. Figure 3 right), noting that she/he grew up with aircraft noise and does not mind it (and even finds it boring without occurrence of flyovers over longer periods of time).
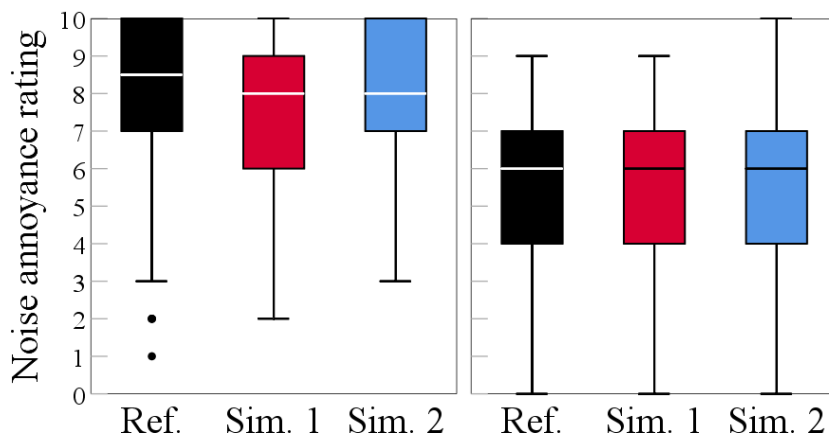

Figure 3: Experimental Part III: Box-and-whisker plots of the individual annoyance ratings of the reference (Ref.) and Simulations 1 and 2 (Sim. 1 and Sim. 2) for the original (left) and the attenuated stimuli (right, additional 15 dB attenuation).

In summary, the psychoacoustic validation revealed that while the auralizations are discriminable from recordings in direct comparisons, they yield similar (and statistically non-significantly different) annoyance ratings. Thus, the proposed methodology using the above simulation–auralization chain can be used for perception-based evaluation of future compared to current aircraft, which is explored in the next section.

3.3 Outlook: Comparison of a future concept with a current aircraft

In the follow-up main study to the validation experiments, auralizations of advanced aircraft configurations (blended wing body) for the year 2050 with possible low-noise technologies (LNT) developed within ARTEM (again based on the two independent simulations) are compared to auralizations (also syntheses, using the same tools) of current aircraft of similar range and mission, namely, a long range tube-and-wing jet aircraft. Figure 4 compares level-time histories of Zwicker loudness [10] for a departure of a current and a future aircraft. Besides very different source directivities of the two aircraft, the future aircraft is also substantially less loud throughout the whole overflight (maximum loudness reduction by a factor of about 4–6 depending on the simulation, cf. Figure 4). Pronounced differences in annoyance ratings are therefore expected. Psychoacoustic laboratory experiments on this were performed and are currently being analyzed.
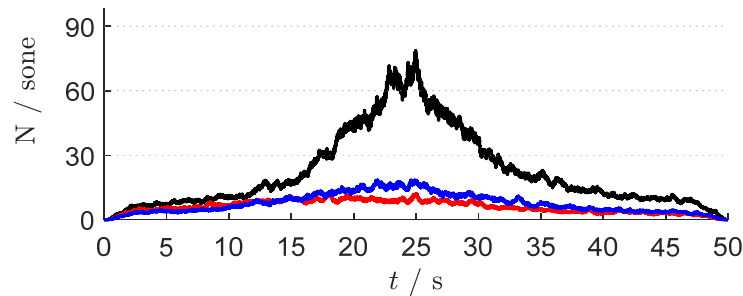

Figure 4: Time-histories of loudness (N) of the departing current (black curve) and future blended wing body aircraft (colored curves) with additional low-noise technologies (LNT) at 9 km distance from the brake release point.

4. CONCLUSIONS

Within the European research project ARTEM, innovative aircraft noise reduction technologies are developed. Using aircraft flyover auralization in laboratory listening experiments, such future technologies are assessed using a perception-based evaluation. To assess the reliability of such an evaluation, the simulation chain was systematically and rigorously validated. The validation experiments revealed that while the sound syntheses are discernable from recordings in direct comparisons, they yield similar (and non-significantly different) annoyance ratings. Thus, the proposed methodological approach (simulation–auralization chain) based on parametric prediction tools such as CARMEN and PANAM combined with the auralization tool AURAFONE can be used for perception-based evaluation of future compared to today's aircraft technologies. Indeed, first results show that advanced aircraft configurations with possible low-noise technologies developed within ARTEM are substantially less loud (and thus presumably less annoying) than current aircraft.

5. ACKNOWLEDGEMENTS

6. REFERENCES

1. EEA: Environmental noise in Europe — 2020. EEA report No 22/2019. European Environment Agency (EEA), Copenhagen, Denmark, 2020, URL: https://www.eea.europa.eu/publications/environmental-noise-in-europe.

2. ICAO: Guidance on the Balanced Approach to Aircraft Noise Management. Doc 9829 AN/451. 2nd ed. International Civil Aviation Organization (ICAO), Montréal, Canada, 2008.

3. IATA: Vision 2050. International Air Transport Association, Singapore, 2011, URL: https://www.iata.org/contentassets/bccae1c5a24e43759607a5fd8f44770b/vision-2050.pdf.

4. ICAO: Annex 16 to the Convention on International Civil Aviation, Environmental Protection, Volume I, Aircraft Noise. 8th Edition, July 2017. International Standards and Recommended Practices. International Civil Aviation Organization (ICAO), Montréal, Canada, 2017.

5. S. A. Rizzi: Toward reduced aircraft community noise impact via a perception-influenced design approach. In: Proceedings of the Inter-Noise 2016, 45th International Congress and Exposition on Noise Control Engineering, Hamburg, Germany, 2016.

6. M. Vorländer: Auralization. Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality. 2nd ed. Springer Nature Switzerland AG, ASA Press, Cham, Switzerland, 2020.

7. R. Pieren, L. Bertsch, D. Lauper, B. Schäffer: Improving future low-noise aircraft technologies using experimental perception-based evaluation of synthetic flyovers. Science of the Total Environment **692** (2019) 68–81.

8. P. Malbéqui, Y. Rozenberg, J. Bulté: Aircraft noise modelling and assessment in the IESTA program. In: Proceedings of the Inter-Noise 2011, 40th International Congress and Exposition on Noise Control Engineering, Osaka, Japan, 2011.

9. L. Bertsch, F. Wolters, W. Heinze, M. Pott-Pollenske, J. Blinstrub: System noise assessment of a tube-and-wing aircraft with geared turbofan engines. Journal of Aircraft **56** (2019) 1577-1596.

10. ISO: ISO 532-1. Acoustics — Methods for calculating loudness — Part 1: Zwicker method. International Standard. International Organisation for Standardization (ISO): Geneva, Switzerland, 2017.

11. L. Bertsch, L. Sanders, R. H. Thomas, I. LeGriffon, J. C. June, I. A. Clark, M. Lorteau: Comparative assessment of aircraft system noise simulation tools. Journal of Aircraft **58** (2021) 867-884.

12. L. Bertsch, G. Looye, E. Anton, S. Schwanke: Flyover noise measurements of a spiraling noise abatement approach procedure. Journal of Aircraft **48** (2011) 436-448.

13. R. Pieren, K. Heutschi, M. Müller, M. Manyoky, K. Eggenschwiler: Auralization of wind turbine noise: emission synthesis. Acta Acustica united with Acustica **100** (2014) 25-33.

14. C. Zellmann, B. Schäffer, J. M. Wunderli, U. Isermann, C. O. Paschereit: Aircraft noise emission model accounting for aircraft flight parameters. Journal of Aircraft **55** (2018) 682-695.

15. I. Le Griffon, L. Bertsch, F. Centracchio, D. Weintraub: Flyover noise evaluation of low-noise technologies applied to a blended wing body aircraft. In: Proceedings of the Inter-Noise 2022, 51st International Congress and Exposition on Noise Control Engineering, Glasgow, UK, 2022.

16. R. Pieren, D. Linke: Auralization of aircraft flyovers with turbulence-induced coherence loss in ground effect. Journal of the Acoustical Society of America **151** (2022) 2453-2460.

17. A. Taghipour, R. Pieren, B. Schäffer: Short-term annoyance reactions to civil helicopter and propeller-driven aircraft noise: a laboratory experiment. Journal of the Acoustical Society of America **145** (2019) 956–967.

18. ISO: ISO/TS 15666. Technical Specification: Acoustics — Assessment of Noise Annoyance by Means of Social and Socio-Acoustic Surveys. 2nd edition 2021-05. International Organisation for Standardization (ISO): Geneva, Switzerland, 2021.

19. B. Schäffer, S. J. Schlittmeier, R. Pieren, K. Heutschi, M. Brink, R. Graf, J. Hellbrück: Short-term annoyance reactions to stationary and time-varying wind turbine and road traffic noise: a laboratory study. Journal of the Acoustical Society of America **139** (2016) 2949–2963.

20. B. Griefahn, A. Marks, T. Gjestland, A. Preis: Annoyance and noise sensitivity in urban areas. In: Proceedings of the 19th International Congress on Acoustics (ICA), Madrid, Spain, 2007.