Metrics for assessing the perception of drone noise

Torija Martinez, AJ and Li, Z

10.48465/fa.2020.0018

METRICS FOR ASSESSING THE PERCEPTION OF DRONE NOISE

A.J. Torija Martinez[1,3]   Z. Li[2]

[1] Acoustics Research Centre, University of Salford, Manchester, UK

[2] Department of Architecture, Zhejiang University of Science & Technology, Hangzhou, China

[3] Institute of Sound and Vibration Research, University of Southampton, Southampton, UK

A.J.TorijaMartinez@salford.ac.uk

ABSTRACT

Unmanned Aerial Vehicle (UAV) technology is rapidly advancing, and therefore, the potential for UAV use seems almost unlimited at this stage. Diverse UAV stakeholders are currently exploring the feasibility of different UAV applications for monitoring, intervention to improve or support public services and parcel delivery. It seems quite likely that, in a short while, communities in urban areas will be inundated with a new source of noise due to UAV operations that they had not before encountered. Noise has been suggested as one of the major barriers of UAVs to public acceptance, and therefore, for the expansion of the sector. The noise of UAVs does not resemble the noise of contemporary aircraft (or any other transportation noise), which leads to an important uncertainty in the prediction of the resultant perception of UAV noise. Previous research has suggested that contemporary noise metrics are unable to account for the qualitative aspects of the particular features of UAV noise. Based on a previous psychoacoustic characterisation of a small fixed-pitch quadcopter, this paper presents the results of a psychoacoustic experiment as a first approach for the development of metrics optimised for UAV noise. Preliminary results suggest that a combined metric including Tonality and Loudness-Sharpness interaction is able to account for the perceptual features of UAV noise.

1. INTRODUCTION

The noise generated by drones does not resemble (qualitatively) the noise of contemporary aircraft, which introduces an unknown factor into the prediction of the noise annoyance. The noise measured for a number of representative small multi-copters is dominated by multiple tones at harmonics of the blade passage frequency (BPF) [1]. Torija et al [2] measured the noise generated by a small quadcopter, the main findings were: (1) the frequency spectra is dominated by a series of harmonic complex tones spaced evenly across the mid-to-high frequency region with relatively event sound levels. (2) The operation of the electric motors generates an important source of noise in the form of high frequency tones.

Christian and Cabell [3] investigated the use of a set of contemporary noise metrics (i.e. Sound Exposure Level –SEL–, Effective Perceived Noise Level –EPNL– and Zwicker loudness) for predicting drone noise annoyance. Christian and Cabell [3] found that none of these metrics are able to appropriately account for the extra annoyance caused by drones as compared to road vehicles, and suggested that novel metrics are required accounting for the quantitative and qualitative aspects of drone noise. A publication of the US Federal Aviation Administration [4] states that existing noise certification methods are not optimal and may not represent the best methods for drone certification, and that noise metrics are needed to better assess subjective response to drone noise.

Among the variety of metrics for transportation noise, only the EPNL accounts for the presence of pure tones. As the primary metric used for aircraft noise assessment, the EPNL was developed to account for the subjective response to a combination of broadband noise and a tonal component (i.e. the BPF of the engine's fan). However, a recent study [5] found that EPNL is unable to appropriately assess the subjective response to aircraft noise with high content in complex tones, and that a more sophisticated tonality method (i.e. Aures tonality) allows a significant improvement over EPNL for such a purpose.

This paper presents the results of a series of listening experiments were participants reported their preference for a number of audio samples of drone and aircraft flyovers, and road vehicle pass-bys. The objective of this research is twofold: (1) quantify and discuss the differences in preference between a small quadcopter and a series of conventional aircraft and road vehicles; and (2) investigate the psychoacoustic metrics more likely to aid the assessment of subjective response to drone noise.

2. MATERIAL AND METHODS FOR PSYCHOACOUSTIC TESTING

2.1 Audio samples

Altogether 34 audio samples were used in the psychoacoustic tests. Eight audio samples of aircraft takeoffs of 2 aircraft types were selected: 4 × A320 (engine CFM56-5) and 4 × A320neo (engine PW1127G). Two extra audio samples of aircraft takeoffs, Boeing 767 (GE90-92B engine) and Boeing 787 (Trent1000 engine), were used as reference sounds, for the preference rating procedure (see below). These audio samples were recorded with a microphone placed at approximately 900 m from the end of the south runway of London Heathrow airport (UK). The recordings contain aircraft passing over the measurement point at a height of approximately 435 m.

Eight audio samples of road vehicles passing-by (single vehicle) were used: 4 × car and 4 × motorbike/mopeds. These road vehicle pass-bys were recorded with a microphone placed at 3.5 m from the edge of the roadway of a busy road in the city of Southampton (UK).

Four audio samples of helicopter flyovers were obtained from 'https://stabserv.larc.nasa.gov/flyover/'. One of the audio samples is an original recording of an AS350 helicopter flyover. The other 3 audio samples are auralisations of an AS350 flyover: audios S7 and S8 in [6] and audio 8 in [7].

Eight audio samples of a series of straight-and-level flyovers of a small quadcopter made were used. The measurements were made in an open field, with the quadcopter (DJI Phantom 3) flying at 2 altitudes above ground level (A = 1 and 2 m), with 2 lateral distances between the microphone and the flight track (L = 0 and 5 m) and with 2 extra-payload conditions (P = 0, 434 and 656 g).

Four audio samples of a quadcopter operated at full power with and without each of its rotor blades were used to investigate the perception of the interaction effects between rotors. These audio samples were recorded with the quadcopter fixed to a stand, and the microphone positioned at 0.96 m away from the quadcopter. Full description of the measurement setup can be found at [2].

As the objective of this research was to understand the perceptual effects of the different frequency composition of drones compared to traditional air and road vehicles, all sounds were normalized to a L_{Aeq} of 65 dB(A). Table 1 shows the list of audio samples used in this research.

Sample ID	Description	EPNL (EPNdB)
A-01	A320 (CFM56-5)	82.2
A-02	A320 (CFM56-5)	81.7
A-03	A320 (CFM56-5)	83.9
A-04	A320 (CFM56-5)	80.5
A-05	A320neo (PW1127G)	83.6
A-06	A320neo (PW1127G)	85.4
A-07	A320neo (PW1127G)	82.9
A-08	A320neo (PW1127G)	83.4
Ref-01	B767 (GE90-92B)	82.4
Ref-02	B787 (Trent1000)	82.4
H-01	AS350 recording	82.9
H-02	AS350 auralisation [6]	83.1
H-03	AS350 auralisation [6]	83.8
H-04	AS350 auralisation [7]	84.6
Q-01	FO-A1-L5-P0	83.1
Q-03	FO-A1-L5-P434	81.8
Q-05	FO-A2-L0-P0	79.8
Q-08	FO-A2-L0-P434	79.5
Q-10	FO-A2-L0-P656	80.9
Q-11	FO-A2-L5-P0	83.9
Q-13	FO-A2-L5-P434	81.7
Q-16	FO-A2-L5-P656	82.0
Q-24	Motor alone	81.8
Q-26	Motor + 1 rotor blade	82.3
Q-28	Motor + 2 rotor blades	82.3
Q-30	Motor + 4 rotor blades	80.0
R-01	Car	80.7
R-02	Car	78.9
R-03	Car	83.5
R-04	Car	81.5
R-05	Moped	82.6
R-06	Moped	81.8
R-07	Motorbike	81.5
R-08	Motorbike	84.1

Table 1. List of audio samples used in psychoacoustic tests.

2.2 Experimental setup

The equipment used for the psychoacoustic tests consisted of a desktop computer (Intel Core i7-2600 CPU @3.40GHz, 16.0 GB RAM, 64-bit Windows 10 Operating System), a USB DAC/headphone amplifier (Audioquest, DragonFly Red v1.2), a pair of open back headphones (AKG K-501). The test was entirely automated via a bespoke MatLab code. The volume level on the laptop was always set to maximum, with MatLab controlling the playback volume to ensure consistency (the reproduced sound levels were not altered after calibration). The tests were carried out in a small anechoic chamber at the Institute of Sound and Vibration Research. The background sound level in this chamber was 15.1 dBA.

2.3 Participants

Thirty participants (16 males and 14 females) with normal hearing ability took part of the psychoacoustic tests. The average age of the participants was 30.5 ± 9.2 years old. A thank you gift of £10 for taking part was used to incentivize participation in the psychoacoustic tests. This experiment was approved by the Ethics and Research committee of the University of Southampton.

2.4 Experimental procedure

In this psychoacoustic test, the participants were asked to rank in order of preference a series of audio samples presented, following a methodology developed by Torija et al. [5]. Eight sets of six audio samples each were created as shown in Table 2.

Set	Audio Samples
S1	Ref-01, Ref-02, A-05, H-01, Q-01, R-05
S2	Ref-01, Ref-02, A-01, H-03, Q-03, R-07
S3	Ref-01, Ref-02, A-02, Q-05, Q-10, R-06
S4	Ref-01, Ref-02, A-06, H-02, H-04, Q-08
S5	Ref-01, Ref-02, A-03, Q-11, Q-16, R-04
S6	Ref-01, Ref-02, A-04, A-07, R-01, R-03
S7	Ref-01, Ref-02, A-08, Q-13, R-02, R-08
S8	Ref-01, Ref-02, Q-24, Q-26, Q-28, Q-30

Table 2. Description of eight sets of six audio samples.

Each set of stimuli comprised 4 test audio samples, and 2 reference audio samples (making six samples in total per set). In total 48 stimuli were ordered by each participant. The 2 reference audio samples selected were 2 types of aircraft with different spectral content and similar loudness. Each set of stimuli comprised audio samples of an aircraft takeoff, helicopter flyover, quadcopter flyover and/or road vehicle pass-by. The aim was to investigate the relative preference magnitude between the type of vehicles, using a 2 reference preference ordering method [5]. In some sets of stimuli, 2 audio samples of the same type of vehicle were included to investigate specific conditions. For instance, the set S3 contained 2 quadcopter audio samples to investigate the noise perception with different payloads (i.e. changes in power). Moreover, the set S8 contained the audio samples recorded at the anechoic chamber with the quadcopter at a fixed position (see above) to investigate the difference in noise perception with a varying number of rotor blades in operation.

This preference ordering method was used to avoid the inherent variation between participants interpretation of a scoring system, and to allow an easier task for the average listener. The 2 reference preference ordering method was selected as it allows a more dynamic assignment of preference for each audio sample in relation to the reference samples.

In each of the eight sets of stimuli, the participants ranked the six audio samples from most preferable to least preferable. During the ranking process, the participants were allowed to listen to each individual audio sample as many times as they required until the final order by preference was decided. Once the final order of preference was confirmed they were able to continue with the test. No specific order of presentation was suggested to the participants, so they could start listening to the six audio samples in the order they wanted. Overall, the participants required between about 40 min to complete the psychoacoustic test.

2.5 Data analysis

The order of preference reported by the participants was transformed into a numerical scale, hereinafter called Preference Rating, using a procedure develop by Torija et al. [5]. An arbitrary magnitude of preference of 60 and 40 was set for the reference audio samples Ref-01 and Ref-02. The PR calculation assumes that the preference magnitude depends not only on the ranking of each sample but the relative spacing of the reference samples. Depending on the order of these 2 references within the set of stimuli, the PR for each audio sample is assigned on the basis of equal spacing using the arbitrary magnitude of the reference samples (See Fig 7 in [5]).

The Sound Quality (SQ) metrics Loudness, Sharpness, Roughness, Fluctuation Strength and Tonality were calculated for each audio sample using the software ArtemiS (Head Acoustics). The method used for Loudness calculations was ISO 532-B. Aures method (with ISO 532-B method set for loudness) was implemented for Sharpness calculations. Terhardt and Aures method [8] was used for Tonality calculations. Default methods in ArtemiS software were used for Roughness and Fluctuation Strength calculations. EPNL was also calculated for each audio sample using an in-house code, developed according to BS 5727:1979 (Method for describing aircraft noise heard on the ground).

The contribution of each SQ metric to PR was assessed using a multilevel modeling approach. In multilevel linear models, regression parameters vary randomly across participants, and therefore are a suitable approach to consider individual responses. All the statistical analyses were carried out with the statistical package IBM SPSS Statistics 25.

3. RESULTS

3.1 Quadcopter vs. Aircraft and Road Vehicles

The audio samples of helicopters were not included in the analysis for this paper. The main reason is the 3 out of the 4 audio samples presented were auralisations, and some participants reported concerns about their realism.

Fig. 1 shows the 5th percentiles of Loudness for each vehicle category. Even though all audio files were set to 65 dB(A), significant differences in Loudness are observed between the aircraft sounds and both the quadcopter and road vehicles sounds. In aircraft audio samples, high frequencies are highly attenuated due atmospheric absorption in the long-range propagation, and therefore low-to-mid frequencies are dominant. This is shown in the lower Loudness values (compared to the other two vehicle categories). With regard to the quadcopter and road vehicle audio samples, the distance between source and microphone was smaller than 6 m, and therefore atmospheric absorption is negligible. The high frequency content in road vehicle sounds, and especially quadcopter sounds, is significantly higher than in aircraft sounds. In Fig. 2, it is shown that the 5th percentiles of Sharpness of the quadcopter audio samples doubles the 5th percentiles of Sharpness of the aircraft audio samples. This high value of Sharpness in the quadcopter audio samples is due to the presence of high frequency tones generated by electric motors [2].

No significant differences in the 5th percentiles of Roughness (Fig. 3) are observed between the audio samples of the three vehicle categories. Fig. 4 shows that the 5th percentiles of Fluctuation Strength of the aircraft audio samples are significantly higher than the ones of both quadcopter and road vehicle audio samples. This might be due to the effects of atmospheric turbulence and wind during the sound propagation from aircraft flyover to microphone.

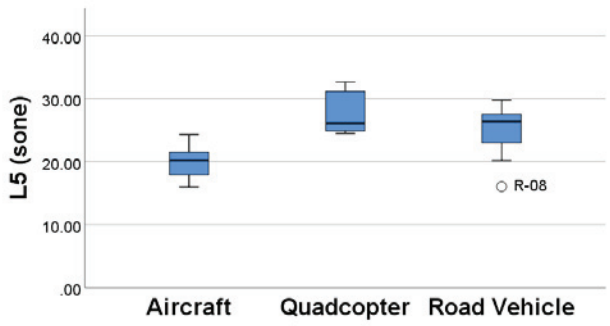

Figure 1. Distribution of 5th percentiles of Loudness values (ISO 532-B) for each vehicle category.

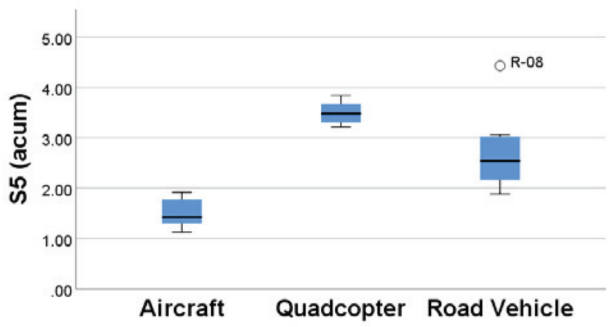

Figure 2. Distribution of 5th percentiles of Sharpness values (ISO 532-B / Aures) for each vehicle category.

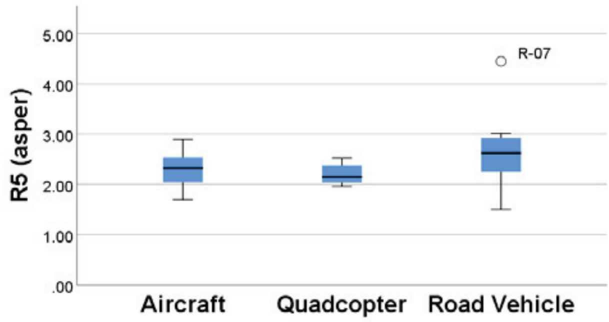

Figure 3. Distribution of 5th percentiles of Roughness values for each vehicle category.

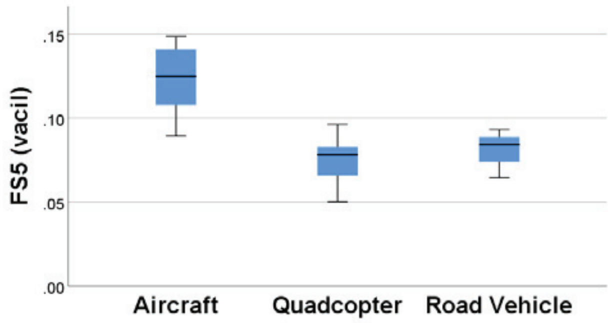

Figure 4. Distribution of 5th percentiles of Fluctuation Strength values for each vehicle category.

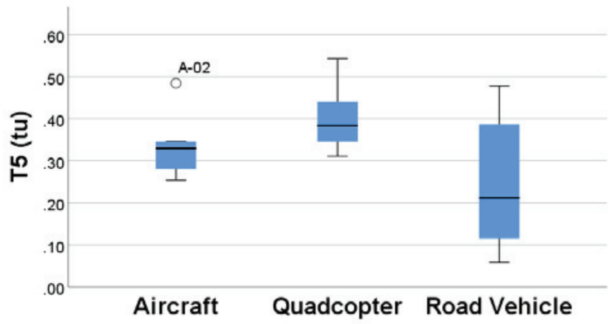

Figure 5. Distribution of 5th percentiles of Tonality values (Terhardt and Aures) for each vehicle category.

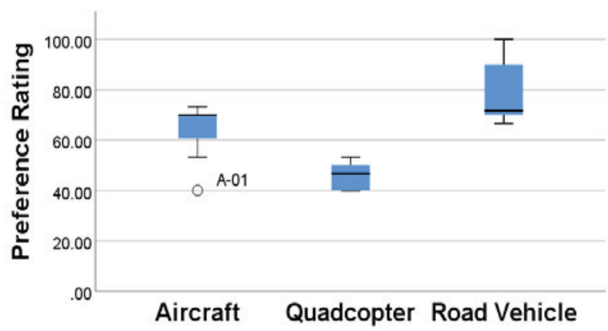

Figure 6. Distribution of Preference Rating values for each vehicle category.

Fig. 5 shows the 5th percentiles of Tonality for each vehicle category. The Tonality of quadcopter audio samples is higher than the Tonality of both aircraft and road vehicles audio samples. The dispersion of Tonality values in the road vehicle category is because audio samples of cars (low Tonality) and motorbikes/mopeds (high Tonality) are included.

Fig. 6 shows the Preference Rating values derived from participants responses, as described above [5], for each vehicle category. Despite all sounds were normalized to a L_{Aeq} of 65 dB(A), the Preference Rating of the quadcopter audio samples is substantially lower than the Preference Rating of both aircraft and road vehicle audio samples. Based on median values, the preference of the quadcopter audio samples is 33% and 35% lower than the preference of the aircraft and road vehicle audio samples respectively. In this paper, it is assumed that these differences in preference might be attributable to specific features of quadcopter sounds, as described by Loudness, Sharpness and Tonality metrics.

3.2 Contribution of Psychoacoustic Factor to Preference Ratings for the Quadcopter

The importance of each psychoacoustic metric to the Preference Rating of the quadcopter audio samples was evaluated using a "one-off" approach. In this approach, the contribution of each metric was assessed based on model accuracy (R^2) when removing it from the analysis [9]. In the multilevel linear model implemented, the intercept was fixed for all participants, and psychoacoustic factors under study were set to vary randomly across participants (i.e. random effects).

As shown in Table 3, only the contributions of Loudness and Tonality to the Preference Rating estimation are statistically significant, and all the other factors were found redundant. Based on the reductions in R^2 shown in Table 3, Tonality is the psychoacoustic factor with highest contribution for estimating the Preference Rating of quadcopter audio samples.

Psychoacoustic Metric	Statistical Significance	Reduction in R^2
Loudness (L5)	0.03	0.01
Sharpness (S5)	0.00*	0.00
Roughness (R5)	0.00*	0.00
Fluctuation Strength (FS5)	0.00*	0.00
Tonality (T5)	0.32	0.03

*Redundant parameter

Table 3. Reduction in conditional R^2 when subtracting individual psychoacoustic factors from the multilevel linear regression model for estimating the Preference Rating.

Interaction between Psychoacoustic Metrics	Statistical Significance	Reduction in R^2
L5-S5	0.43	0.02
L5-R5	0.00*	0.00
L5-FS5	0.00*	0.00
L5-T5	0.02	0.01
S5-R5	0.00*	0.00
S5-FS5	0.00*	0.00
S5-T5	0.00*	0.00
R5-FS5	0.00*	0.00
R5-T5	0.00*	0.00
FS5-T5	0.00*	0.00

*Redundant parameter

Table 4. Reduction in conditional R^2 when subtracting individual interactions between psychoacoustic factors from the multilevel linear regression model for estimating the Preference Rating.

Table 4 shows the contribution of each interaction between psychoacoustic metrics for estimating the Preference Rating of the quadcopter audio samples. As shown in Table 4, only the contributions of the Loudness-Sharpness interaction and Loudness-Tonality interaction are statistically significant. Among all the possible interactions between psychoacoustic metrics, the Loudness-Sharpness interaction is the highest contributor to the Preference Rating estimation of quadcopter audio samples.

A multilevel linear model with fixed intercept and Tonality and Loudness-Sharpness interaction as random effects estimates the Preference Rating for the quadcopter audio samples with an $R^2 = 0.69$.

Exploratory analysis during the implementation of the multilevel linear modeling, with and without the inclusion of random effects, suggests the participant as a significant factor in determining the Preference Rating of the quadcopter audio samples.

3.3 Perception of Rotors Interaction Effects

Table 5 shows the Preference Rating of the audio samples of the quadcopter measured in a fixed position at an anechoic chamber with only the motor operating (Q-24), and with the motor and 1 (Q-26), 2 (Q-28) and 4 (Q-30) rotor blades operating.

As shown in Table 5, even with the same L_{Aeq} (65 dB(A)), the audio samples with the rotor blades operating were reported significantly less preferable than the one with only the motor operating. Compared to 1 and 2 rotor blades operating, the audio sample with the 4 rotors operating is substantially less preferable. This finding might be due to the interaction effects between rotor blades, as discussed in [2].

Audio Sample	Mean	95% Confidence Interval	
		Lower Bound	Upper Bound
Q-24	55.00	38.54	71.46
Q-26	29.00	11.20	46.80
Q-28	37.33	21.74	52.93
Q-30	23.33	3.25	43.41

Table 5. Mean value and 95% Confidence Interval of the mean for quadcopter audio samples with varying number of rotor blades in operation.

4. CONCLUSIONS

This paper presents the results of a psychoacoustic experiment, where a number of participants were asked to rank in order of preference a series of quadcopter, aircraft and road vehicle audio samples. For the specific audio samples tested in this psychoacoustic experiment, the most significant results are:

- Even with the same L_{Aeq}, the preference of the quadcopter audio samples was 33% and 35% lower than the preference of aircraft and road vehicle audio samples respectively.
- Based on a multilevel linear modeling, Tonality and Loudness-Sharpness interaction are the two main psychoacoustic factors determining the Preference Rating of the quadcopter audio samples.
- The interaction effect between rotor blades seems to influence the perception of quadcopter sounds. Further research will be carried out to investigate in-depth the perception of rotor interaction effects.

5. REFERENCES

[1] R. Cabell, R. McSwain, F. Grosveld: "Measured noise from small unmanned aerial vehicles," *Proceedings of the NOISE-CON 2016*, 13-15 Jun. 2016, Providence, RI, United States.

[2] A.J. Torija, R.H. Self, J.L.T. Lawrence: "Psychoacoustic Characterisation of a Small Fixed-pitch Quadcopter," *Proceedings of Inter-Noise 2019"*, paper 1423, 16-19 Jun. 2019, Madrid, Spain.

[3] A.W. Christian, R. Cabell: "Initial investigation into the psychoacoustic properties of small unmanned aerial system noise," *Proceedings of the 23rd AIAA/CEAS Aeroacoustics Conference, AIAA AVIATION Forum, (AIAA 2017-4051).*

[4] D.A. Senzig, M. Marsan: "UAS Noise Certification," *In proceedings of Inter-Noise 2018*, paper in18_1855, 2018, Chicago, United States.

[5] A.J. Torija, S. Roberts, R. Woodward, I.H. Flindell, A.R. McKenzie, R.H. Self: "On the assessment of subjective response to tonal content of contemporary aircraft noise," *Applied Acoustics*, 146, 190-203, 2019.

[6] N.M. Pera, S.A. Rizzi, S. Krishnamurthy, C.R. Fuller, A. Christian: "A Recording-Based Method for Auralization of Rotorcraft Flyover Noise," *In proceedings of AIAA SciTech 2018*, AIAA-2019-0267, Kissimmee, FL, USA, 8-12 Jan 2018.

[7] S. Krishnamurthy, S.A. Rizzi: "Auralization of Amplitude Modulated Helicopter Flyover Noise," *In proceedings of AIAA SciTech 2019,* AIAA-2019-2087, San Diego, CA, USA, 7-11 January 2019.

[8] W. Aures: “"Berechnungsverfahren für den sensorischen Wohlklang beliebiger challsignale (Procedure for calculating the sensory euphony of arbitrary sound signals)," *Acta Acustica united with Acustica*, 59(2), 130-141, 1985.

[9] M. Boucher, S. Krishnamurthy, A. Christian, S. Rizzi: "Sound quality metric indicators of rotorcraft noise annoyance using multilevel regression analysis," *In Proceedings of the 177th Meeting of the Acoustical Society of America*, 145, 1899, 23 April 2019.