forum acusticum 2026
8-12 September · Graz, Austria

JUST NOTICEABLE DIFFERENCE AND ABSOLUTE THRESHOLDS OF PSYCHOACOUSTIC ROUGHNESS FOR SIMPLE REFERENCE AND COMPLEX DRONE SIGNALS

Felix Hochbaum[1], André Fiebig[1]

[1] *Department of Engineering Acoustics, Technische Universität Berlin, Einsteinufer 25, 10587 Berlin, Germany*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

The psychoacoustic metric of roughness is key to many acoustic evaluations. Although several calculation methods are available, each method yields slightly different roughness values for different signal and modulation types. In addition, the practical interpretation of roughness values is often unclear due to the limited knowledge of absolute and just-noticeable difference thresholds, especially for complex technical sounds, which are typically subject of roughness analysis. To broaden the empirical basis on thresholds of roughness, a listening test was designed and conducted. Using a constant-stimuli procedure, thresholds were determined for both simple reference signals, including amplitude-modulated pure tones and broadband noise, and more complex synthesized drone sounds. Thresholds were derived for three commonly used roughness calculation methods (ECMA 418-2, Zwicker & Fastl and Daniel & Weber). Although threshold values depended on signal type and calculation method, low roughness values were associated with low thresholds, whereas higher roughness levels resulted in higher thresholds.

Keywords: *psychoacoustic roughness, threshold, JND, psychoacoustic parameters*

1. Introduction

Psychoacoustic roughness addresses the auditory perception of rapid modulations (for an overview see [1]) and is commonly used to characterize a wide variety of signal types, including noise of engines, household appliances or human voice. While several calculation methods of roughness are available, the practical interpretation of roughness values remains difficult due to the still limited knowledge of absolute thresholds and just noticeable differences (JNDs). At what point is a sound perceived as rough? Which differences in calculated roughness values are relevant and perceptible? Existing thresholds are either model-specific or derived from idealized signals such as amplitude-modulated pure tones, whereas practical roughness assessment usually concerns complex technical sounds with modulations.

Acoustic emissions of unmanned aerial vehicles or drones are receiving increasing attention in psychoacoustic research and are characterized, among other properties, by complex modulation patterns, including stochastic amplitude and frequency modulations of tonal components and periodic modulation of broadband components at the blade-passing frequencies (BPFs) [2], [3]. Auditory roughness as percept is particularly prominent in amplitude- or frequency-modulated sounds and depends on the temporal and spectral resolution of the auditory system [1], [4]. Although roughness has mostly been studied using periodically modulated signals, it may also arise from irregular envelope fluctuations, such that even unmodulated noise can evoke a rough auditory sensation [5]. Roughness and modulation difference thresholds have been studied extensively since the mid 20th century, with reported thresholds for amplitude modulated pure tones ranging from as little as 8% to about 20% relative change in modulation depth (m), depending on modulation conditions (mainly carrier and modulation frequency, f_c and f_{mod}) [6], [7], [8], while Suchowersky reported a threshold of $\Delta m = 0.1$ [9]. Already in 1974, Vogel recognized that roughness should be studied using more complex sounds than amplitude-modulated pure tones, showing that partial masking reduced roughness and that a broadband 20-tone complex elicited substantially greater roughness than a 1-kHz pure tone [10]. For narrowband noises, Fleischer found absolute thresholds ranging from about $m = 0.1$ to $m = 0.4$ that increased with modulation frequency, while difference thresholds remained around 20-40% across conditions [11], [12]. One important finding, re-

Corresponding author: *felix.hochbaum@tu-berlin.de.*

garding the absolute threshold of narrow band noises, is that it cannot be lower then that of amplitude modulated pure tones [12]. Viemeister (1979, as cited in [13]) reported a modulation-detection threshold of approximately $m = 0.1$ for sinusoidally amplitude-modulated wideband noise (f_m = 100 Hz), higher than known detection thresholds for pure tones [13]. Zwicker and Fastl relate a just-noticeable modulation change of 10% to a roughness change of about 17% and state, for a 1 kHz tone amplitude-modulated at f_{mod} = 70 Hz, an absolute threshold of 0.07 asper [1]. Daniel and Weber [14] refer to these values. More recent work has also addressed roughness JNDs for complex technical sounds. Leutheuser et al. [15] reported modulation-depth difference thresholds of about 10–22% for amplitude-modulated pure tones with additional pink noise, in line with earlier findings [1], [6], [7]. For engine sounds, no uniform roughness JND was identified, instead, perceptibility seemed to depend on the modulation spectrum and the specific signal characteristics [15]. You and Jeon [16] determined a JND for roughness according to the Zwicker & Fastl model with modified refrigerator sounds and determined difference thresholds of 0.04 asper below, and 0.05 asper above the reference stimulus.
To extend the empirical database on thresholds of roughness perception, absolute and difference thresholds were determined for two simple reference signals (pure tone, pink noise), as well as for two synthesized drone signals and one synthesized rotor signal whose roughness was systematically varied. The thresholds were then derived for three popular roughness models: (1) Calculation based on the Sottek Hearing Model as standardized in ECMA 418-2 [17] and in DIN 38455 [18], the model by Zwicker & Fastl [1] as available in MATLAB® Audio Toolbox and (3) the model by Daniel & Weber [14] as available in the SQAT-Toolbox [19].

2. Method

To estimate absolute and just-noticeable difference thresholds for auditory roughness, a listening experiment was conducted using a constant-stimulus two-interval, two-alternative forced-choice (2I-2AFC) paradigm. In each trial, participants selected the rougher of two signals, thereby directly assessing roughness perception.

2.1. Participants

Due to the demanding and time-consuming nature of the experiment, participants were recruited from members of the research group and affiliated labs. A total of N = 10 participants took part in the listening experiment, all stated male gender. The average age was 31.8 years (SD = 6.7). All participants self-reported normal hearing, nine out of ten people claimed to have a trained hearing. All participants gave informed consent to scientific use and publication of the data.

2.2. Stimuli

A selection of simple reference and complex synthesized drone sounds was made. Roughness was systematically manipulated by varying the modulation depth m of the entire signal or signal components, resulting in stimulus sets with finely spaced roughness levels and a monotonic relationship between the (internal) modulation index $(i)m$ and roughness. Stimuli and step sizes were defined a priori. All stimuli had a duration of 2 s and a sampling rate of 48 kHz. To avoid clicking artifacts, 10 ms raised-cosine ramps were applied at stimulus onset and offset. For the listening test, all signals were normalized to the same RMS level and presented at 60 dB SPL.

2.2.1. Simple reference signals

Two sinusoidally amplitude-modulated reference signals were used: a 1-kHz tone and pink noise. For both signals, stimulus sets were generated for modulation depths ranging from $m = 0$ to $m = 1$ in increments of $\Delta m = 0.01$. For each reference signal, three anchor points were selected, from which six higher comparison intervals were presented. The 1-kHz tone was amplitude modulated at $f_{\text{mod}} = 70$ Hz. At full modulation, this signal corresponds to the reference defining 1 asper. In contrast to this definition, however, all stimuli in the present experiment were normalized to the same RMS level regardless of modulation depth, whereas the overall level of the 1 asper reference increases slightly with increasing modulation. Reference anchors were set to $m_{\text{ref}} = 0$, corresponding to the absolute threshold, $m_{\text{ref}} = 0.4$, and $m_{\text{ref}} = 0.8$, from which six comparison intervals were presented in steps of $\Delta m = 0.03$. For the noise reference, pink noise was generated with a fixed seed to ensure deterministic reproduction, band-pass filtered from 100 Hz to 10 kHz using a fifth-order zero-phase forward-backward filtering to minimize phase distortion, and sinusoidally amplitude-modulated at $f_{\text{mod}} = 100$ Hz. Reference anchors were set to $m_{\text{ref}} = 0.0$, 0.4, and 0.76, with six comparison intervals in steps of $\Delta m = 0.04$. Due to the larger step size, the highest anchor was compared up to full modulation ($m = 1.0$).

2.2.2. Synthesized drone signals

Because a synthesis-based approach allows the variation of individual signal properties, two hovering drone signals measured at TU Berlin (X500, UBADRON) and one drone rotor signal from the DREGON dataset [20] were synthesized and systematically varied in roughness. The synthesis followed Dreier et al. [6], with several modifications and extensions, most notably to allow systematic quasi-periodic modulation of the broadband component. The synthesis method can be divided into two main stages: (1) analysis of the reference signal, in which the relevant information is extracted, and (2) signal synthesis, in which a new signal is generated. The synthesis stage itself can further be subdivided into the generation of the tonal components and the broadband component, which are summed in a final step. Most parameters

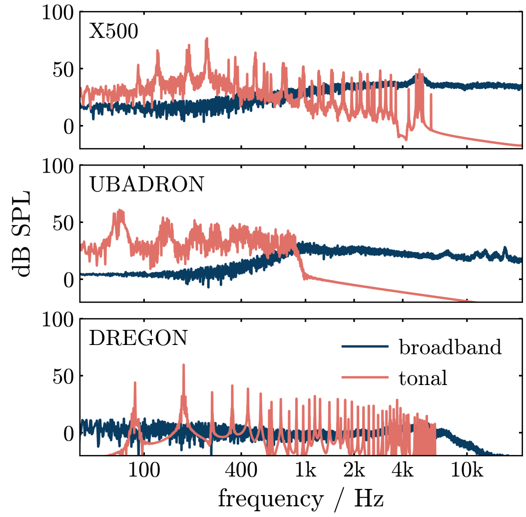

Figure 1: Smoothed FFTs of the tonal and broadband parts of the three synthesized reference drone sounds. The broadband component was quasiperiodically modulated at the blade-passing frequencies (BPFs) and the internal modulation index (im) systematically varied.

were derived from signal analysis, others were chosen heuristically. The goal was to obtain a reference synthesis perceptually close to the original signal while retaining full control over all perceptually relevant parameters. Fixed assumptions and predefined random seeds ensured that repeated synthesis with identical parameters yielded identical signals. Tonal peaks and their amplitudes were extracted from the spectrum of the reference measurement. This information was also used to identify the BPFs and to remove the tonal components from the spectrum. For peak removal, notch-filterbank-based methods, median filtering, and time-domain noise-estimation procedures were applied. The remaining broadband spectrum was modeled using the autoregressive approach to derive the filter coefficients for the synthesis of the noise component as described in Dreier et. al. [2]. Each tonal component was generated iteratively with slight stochastic amplitude and frequency modulation. The instantaneous frequency vectors of the identified BPFs were retained to modulate the broadband component and to couple its modulation rate to the respective instantaneous BPF. Summing all tonal components yielded the tonal part of the synthesis. The broadband component was synthesized by filtering white noise with the derived filter coefficients and restricting it to the relevant frequency range by band-pass filtering. Previous work has shown that the high-frequency broadband component is periodically modulated at the BPF [3], while also exhibiting aperiodic variability [21]. Accordingly, the filtered noise signal was modulated using BPF-specific modulators whose properties were defined by the stored instantaneous frequency vectors, the modulation shape, and the modulation depth (here referred to as the internal modulation index im), thereby coupling the broadband modulation to the BPFs and their stochastic frequency fluctuations. For each BPF considered, a separate noise spectrum was modulated; these components were then summed and normalized to yield the broadband part of the synthesis. A more detailed description of the synthesis approach can be found in our previous work [22]. For the X500 synthesis, two closely spaced BPFs (242 Hz and 249 Hz) were identified in the reference measurement and used to sinusoidally modulate two noise spectra. An internal modulation index of $im_{ref} = 0.05$ was chosen as the anchor, from which six comparison intervals were presented in the listening experiment in steps of $\Delta im = 0.15$. The first two BPFs (65 Hz and 74.5 Hz) and their first harmonics (130 Hz and 149 Hz) were taken into account for the broadband modulation of the UBADRON synthesis. Starting from $im_{ref} = 1$, the threshold was determined as a descending threshold in steps of $\Delta im = -0.08$. In the DREGON rotor synthesis, the broadband spectrum was modulated exclusively by the BPF at approximately 176 Hz, using a square-wave modulator with a duty cycle of 20%. Starting from $im_{ref} = 0.5$, six levels with $\Delta im = 0.05$ were presented.

2.3. Experimental setup

The listening test was conducted in an audiometry cabin at Technische Universität Berlin. A custom-built Graphical User Interface programmed in Python was used to control the test procedure and handle data collection and signal playback. Digital-to-analog conversion was performed by a Fireface UCX II, which fed a Sennheiser HD 600 headphone. Matching the stimuli, the interface ran at a sampling rate of 48 kHz. Listening test playback levels were calibrated using a HEAD acoustics HMS III dummy head measurement system.

2.4. Procedure

A classic two-interval, two-alternative forced-choice (2I-2AFC) discrimination procedure was employed. In each trial, participants were presented with an anchor and a variable comparison stimulus and indicated which interval was perceived as rougher [23]. Presentation order was randomized. Nine signal groups were tested: three 1 kHz tone conditions, three noise conditions, and three synthesized drone sounds. All stimulus sets and trials were interleaved and shuffled. Each set comprised one anchor and six comparison stimuli. With 30 trials per comparison level, this yielded $6 \times 30 = 180$ trials per set and 1620 decisions in total. To reduce fatigue, the experiment was split into two equal sessions of 15 trials per comparison level, completed either on the same or different days. After completing a questionnaire and generating an anonymized participant code, participants were informed about the aim of the study and instructed on the task in the audiometry booth. An instruction screen briefly introduced roughness and presented three sounds for familiarization: two sine tones (500 Hz and 1 kHz) and one pink-noise stimu-

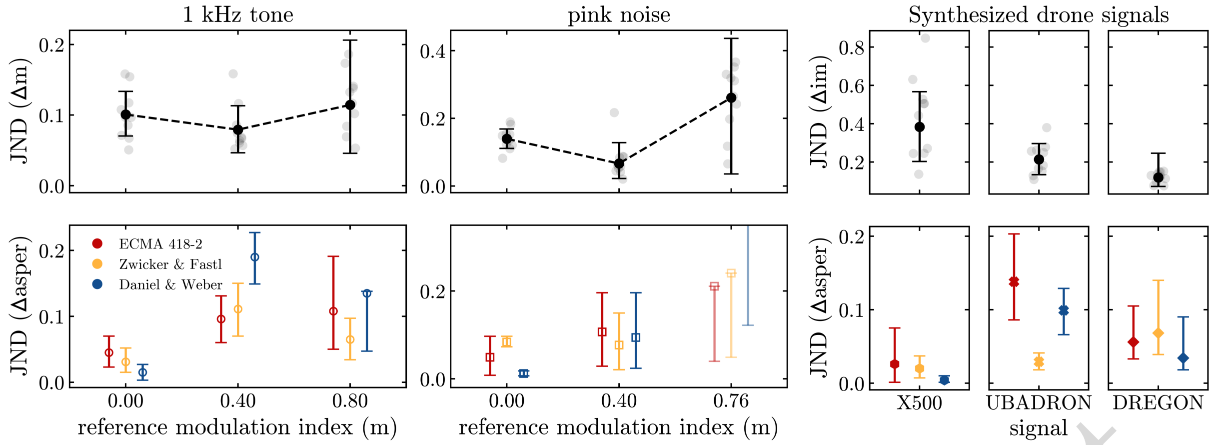

Figure 2: Determined thresholds for all signals and reference levels, expressed as JND of $\Delta(i)m$ (top), and JND of roughness $\Delta asper$, for the three roughness models (bottom), with error bars indicating the 95% HPDIs. Synthesized drone signals are shown ascending roughness of the anchor signals after ECMA 418-2. Reduced opacity indicates invalid threshold estimates.

lus, each in an unmodulated and a modulated version corresponding to 1 asper according to ECMA 418-2. Each trial consisted of two 2 s signals separated by 0.5 s of silence and was presented only once. Participants advanced after each response by clicking "Next Pair" and breaks were allowed at any time. Session duration was approximately 1 h 45 min, corresponding to about 3.5 h total experimental duration per participant.

2.5. Analysis

Listening test data were pooled across the two sessions for each participant analysis was then performed separately for each reference stimulus set (anchor). In constant-stimulus experiments, thresholds are typically estimated by fitting a psychometric function (PF) to binary response data across stimulus intensities. In the present study, stimulus strength was defined as modulation depth m or, for drone synthesis, the internal modulation index im, forming the abscissa of the PF. The ordinate gives the probability of a correct response, with the lower asymptote fixed by the guess rate (γ). The remaining PF parameters are the slope (β), describing the steepness of the function around threshold, lapse rate (λ), and threshold (α), defined as the midpoint between the lower and upper asymptotes [23]. In a 2AFC task, the threshold is 0.75 without lapses and decreases when lapses are included. Since population level thresholds were of primary interest in this study, a hierarchical Bayesian model was chosen. By estimating individual- and group-level parameters simultaneously, such multilevel (or partial-pooling) models have been shown to improve group-threshold estimates compared with no-pooling and complete-pooling approaches [24]. The model used is implemented in the Palemedes MATLAB Toolbox [25] and calls on JAGS [26]. Psychometric functions with a cumulative probit link were fitted using default priors. The guess rate (γ) was constrained to 0.5 to reflect the 50% chance level in the 2-AFC task, whereas all remaining parameters were estimated freely. In the descending-threshold estimation of UBADRON, the x-axis had to be inverted to fit an ascending psychometric function. Population-level thresholds were derived as posterior medians of the group-level parameters and, together with their 95% highest-posterior-density intervals (HPDIs), mapped onto roughness values for the ECMA 418-2, Zwicker & Fastl, and Daniel & Weber models. This was done by calculating the roughness values of the stimuli closest to the estimated threshold and the corresponding HPDI bound. Because all stimuli were generated in modulation steps of $\Delta(i)m = 0.01$, this mapping was straightforward but effectively involved rounding to two decimal places. The roughness single-number values were derived as the 90th percentile (R_{90}) for ECMA 418-2 and as arithmetic means (R_{mean}) for the Zwicker & Fastl and Daniel & Weber methods.

3. Results

Table 1 lists the group-level posterior parameters (threshold, slope, lapse rate), along with the roughness values of the anchor and threshold signals and the corresponding relative change for all signals. Group thresholds as Δm together with their 95% HPDI, individual thresholds, and mapped thresholds as $\Delta asper$ for the three roughness models are shown for all signals in Figure 2. Both reference signals (1 kHz tone and pink noise) showed a similar pattern: JNDs were lowest at the mid-modulation condition ($m_{ref} = 0.4$), slightly higher at threshold, and highest at $m_{ref} = 0.8/0.76$. For both signals, hyperslopes decreased and lapse rates increased with modulation depth, while individual threshold spread and credible intervals were largest at the highest modulation condition. For pink noise, the highest modulation condition ($m_{ref} = 0.76$) produced a ceiling effect, with the estimated threshold ($m = 1.021$) exceeding the tested range ($m_{ref} = 0.76$ to $m = 1.0$ in $\Delta m = 0.04$

Table 1: Results of the listening experiment with reference (internal) modulation index, estimated group-thresholds, slopes and lapse rates as medians of the posterior distribution with 95% highest posterior density intervals and roughness values in *asper* of the reference (anchor) signals and thresholds for all three roughness models including relative changes.

Signal	Reference	Threshold μ_α	Slope μ_β	Lapse μ_λ	ECMA 418-2 $R_{\text{ref}}/R_{\text{thres}}(\%)$	Zwicker & Fastl $R_{\text{ref}}/R_{\text{thres}}(\%)$	Daniel & Weber $R_{\text{ref}}/R_{\text{thres}}(\%)$
1 kHz Tone	$m = 0.00$	0.100 [0.07, 0.13]	1.846 [1.32, 2.47]	0.026 [0, 0.09]	0.000 / 0.045	0.000 / 0.031	0.001 / 0.016
	$m = 0.40$	0.479 [0.45, 0.51]	1.186 [0.94, 1.46]	0.047 [0, 0.09]	0.359 / 0.455 (27%)	0.374 / 0.485 (30%)	0.301 / 0.491 (63%)
	$m = 0.80$	0.914 [0.85, 1.00]	1.103 [0.33, 1.99]	0.134 [0.03, 0.26]	0.809 / 0.917 (13%)	0.821 / 0.886 (8%)	0.862 / 0.997 (16%)
Pink Noise	$m = 0.00$	0.139 [0.11, 0.17]	1.432 [1.22, 1.66]	0.013 [0, 0.05]	0.024 / 0.073 (304%)	0.082 / 0.166 (202%)	0.031 / 0.043 (39%)
	$m = 0.40$	0.465 [0.42, 0.53]	1.038 [0.79, 1.34]	0.037 [0, 0.1]	0.534 / 0.641 (20%)	0.431 / 0.354 (22%)	0.197 / 0.291 (48%)
	$m = 0.76$	1.021 [0.8, 1.2]	0.740 [-0.12, 1.92]	0.178 [0, 0.33]	0.997 / 1.208 (21%)	0.81 / 1.051 (30%)	0.963 / 1.7 (77%)
X500	$im = 0.05$	0.434 [0.25, 0.62]	0.619 [0.41, 0.85]	0.011 [0, 0.04]	0.002 / 0.028 (–)	0.048 / 0.068 (42%)	0.018 / 0.022 (22%)
UBADRON	$im = 1(0)$	0.214 [0.13, 0.3]	0.747 [0.51, 0.99]	0.014 [0, 0.05]	0.383 / 0.245 (-36%)	0.172 / 0.143 (-17%)	0.226 / 0.127 (-44%)
DREGON	$im = 0.5$	0.611 [0.57, 0.71]	1.44 [0.99, 1.8]	0.128 [0, 0.22]	0.570 / 0.626 (10%)	0.216 / 0.284 (31%)	0.094 / 0.128 (36%)

steps). As this implies slight overmodulation, no valid threshold was determined; for the present analysis, roughness was therefore assigned to the highest tested level ($m = 1.0$). Except at mid modulation, where JNDs were close, pink noise appears to produce higher thresholds (in m) than the modulated pure tone.

Thresholds for the synthesized drone signals are less directly interpretable, as the internal modulation index does not correspond to the modulation depth of a fully modulated signal and is barely comparable across syntheses because of the different modulation schemes. What can be said is that thresholds varied widely, ranging from one or two steps above the reference to not being reached even at the largest presented im difference (especially for X500, see Figure 2). DREGON showed a comparatively steep slope, whereas the other two synthesized signals exhibited much lower hyperslopes.

4. Discussion

The similar pattern observed for both reference signals suggests that roughness discrimination is most sensitive at intermediate modulation depths and becomes more difficult as modulation depth increases, with thresholds increasing accordingly. The absolute threshold of the modulated 1 kHz tone, expressed in modulation depth, was slightly higher than some literature values, such as those reported by Kohlrausch [13], but remained below the 0.07 *asper* threshold given by Zwicker and Fastl [1]. Similarly, the modulation-detection threshold of the pink-noise stimulus exceeded the value reported by Viemeister (1979) [13] for comparable noise stimuli. The relative JNDs of modulation depth, amount to 19.8 % and 14.3 % for the pure tone and 16,3% and 34,3% for the pink noise and gather around $\Delta m = 1$. These values fall within the range reported by Leutheuser et al. [15], Suchowerskyj [9], Terhardt [6] and Fleischer [11], [12]. Noise generally showed higher thresholds than the pure tone, in line with Fleischer [12], possibly due to its broadband nature, which also yields higher roughness values, or to the higher modulation frequency of 100 Hz compared to 70 Hz. For both reference signals at $m_{\text{ref}} = 0$ and for the X500 signal, where roughness was close to zero *asper* in all three models, the Δ*asper* thresholds were consistently lower than under conditions with higher modulation or roughness. The thresholds for low roughness ranged from 0.025 to 0.05 for ECMA, 0.02 to 0.08 for Zwicker & Fastl, and 0.004 to 0.012 for Daniel and Weber. Notably, all three models predicted higher baseline roughness for unmodulated pink noise than for the X500 signal. With higher modulation and higher roughness, thresholds altered to around 0.056 to 0.138 (pink noise: 0.21) for ECMA, 0.029 to 0.111 (pink noise: 0.25) for Zwicker & Fastl, and 0.034 to 0.19 (pink noise: 0.737) for Daniel & Weber. Expressed as relative changes in % with respect to roughness as reported in table 1, the JNDs may be larger than sometimes assumed, although no clear general rule can be derived. For UBADRON and DREGON, the higher roughness (compared to X500) also tended to be associated with higher thresholds; however, with the exception of the ECMA threshold for UBADRON, these thresholds remained below that of the reference signals. In the Zwicker & Fastl model, the JNDs of $\Delta 0.02$, $\Delta 0.029$, and $\Delta 0.068$ *asper* were around the value of $\Delta 0.04 - 0.05$ *asper* reported by You and Jeon [16] as JND for a complex signal.

Although signal synthesis proved suitable for targeted signal generation, drones constitute a special case due to their complex modulation and dominant tonal components, which may partially mask modulation. The comparatively high thresholds and interindividual spread may be attributable to both the potentially fatiguing constant-stimuli procedure and the nature of the task, which required evaluation of the roughness percept rather than mere difference detection. Sometimes participants reached threshold at the smallest presented roughness distance, whereas others did so only at the largest. This should be taken into account when interpreting the results, also as no outlier correction was applied.

5. Conclusion

Absolute and difference thresholds of psychoacoustic roughness were assessed in a listening experiment using a 2I-2AFC procedure with simple reference signals and complex synthesized drone sounds that were systematically varied in roughness. Threshold estimates showed some interindividual variability. Furthermore, just-noticeable differences depended on signal type and differed across roughness models. In summary, the results across all models suggest that thresholds can be very low ($\leq \Delta 0.05 asper$) at low roughness levels and tend to be somewhat higher at larger roughness values, without increasing monotonically.

Acknowledgments

The authors would like to thank the HEAD-Genuit Foundation for funding this work (P-16/05-W) and everybody who participated in the experiment.

References

[1] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*, 3rd. Berlin: Springer, 2007.

[2] C. Dreier and M. Vorländer, "Drone auralization model with statistical synthesis of amplitude and frequency modulations," *Acta Acust.*, vol. 8, p. 35, 2024.

[3] W. J. Baars and D. Ragni, "Low-frequency intensity modulation of high-frequency rotor noise," *AIAA J.*, vol. 62, pp. 3374–3390, 2024.

[4] C. Maschke and A. Jakob, "Psychoakustische messtechnik," in *Messtechnik der Akustik*, Springer, 2009, pp. 599–642.

[5] A. Anikin, "Acoustic estimation of voice roughness," *Atten. Percept. Psychophys.*, vol. 87, pp. 1771–1787, 2025.

[6] E. Terhardt, "Über die durch amplitudenmodulierte sinustöne hervorgerufene hörempfindung," *Acustica*, vol. 20, pp. 210–214, 1968.

[7] A. Vogel, "Über den zusammenhang zwischen rauhigkeit und modulationsgrad," *Acustica*, vol. 32, pp. 300–306, 1975.

[8] H. Fleischer, "Schema zur berechnung der hörbarkeitsschwellen von phasenunterschieden," *Biol. Cybern.*, vol. 23, pp. 161–170, 1976.

[9] W. Suchowerskyj, "Beurteilung von unterschieden zwischen aufeinanderfolgenden schallen," *Acustica*, vol. 38, pp. 131–139, 1977.

[10] A. Vogel, "Roughness and its relation to the time pattern of psychoacoustical excitation," in *Facts and Models in Hearing*, E. Zwicker and E. Terhardt, Eds., Springer, 1974, pp. 238–247.

[11] H. Fleischer, "Modulationsschwellen von schmalbandrauschen," *Acustica*, vol. 51, pp. 154–161, 1982.

[12] H. Fleischer, "Psychoakustische kennwerte von amplitudenmoduliertem schmalbandrauschen," Univ. Bundeswehr München, Tech. Rep., 1981.

[13] A. Kohlrausch, R. Fassel, and T. Dau, "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," *J. Acoust. Soc. Am.*, vol. 108, pp. 723–734, 2000.

[14] P. Daniel and R. Weber, "Psychoacoustical roughness: Implementation of an optimized model," *Acta Acust.*, vol. 83, pp. 113–123, 1997.

[15] A. Leutheuser, R. Sottek, and J. Becker-Schweitzer, "Just-noticeable roughness differences of technical sounds," in *Proc. NAG/DAGA 2009*, 2009, pp. 384–387.

[16] J. You and J. Y. Jeon, "Just noticeable differences in sound quality metrics for refrigerator noise," *Noise Control Eng. J.*, vol. 56, pp. 414–423, 2008.

[17] *ECMA-418-2, Psychoacoustic Metrics for ITT Equipment—Part 2, 3rd ed.* 2024.

[18] *DIN 38455:2022-11, Berechnung der Hörempfindung Rauigkeit*, 2022.

[19] G. F. Greco, R. Merino-Martínez, and A. Osses, *SQAT: A sound quality analysis toolbox for MATLAB*, Software, 2025.

[20] M. Strauss, P. Mordel, V. Miguet, and A. Deleforge, "DREGON: Dataset and methods for UAV-embedded sound source localization," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2018, pp. 1–8.

[21] Z. F. Gan, V. Tumelero Valente, K. S. Brentner, and E. Greenwood, "Broadband noise modulation of multirotor aircraft," *J. Acoust. Soc. Am.*, vol. 157, pp. 924–939, 2025.

[22] F. Hochbaum and A. Fiebig, "Untersuchung von schwellen der rauigkeitsempfindung mit einfachen referenzsignalen und komplexen synthetisierten drohnengeräuschen," in *Proc. DAGA 2026*, 2026.

[23] R. Ulrich and D. Vorberg, "Estimating the difference limen in 2AFC tasks: Pitfalls and improved estimators," *Atten. Percept. Psychophys.*, vol. 71, pp. 1219–1227, 2009.

[24] M. A. Boucher, A. W. Christian, M. Rafaelof, K. P. Shepherd, S. A. Rizzi, and J. H. Stephenson, "Comparisons of analysis methods applied to alternative forced choice audibility data," NASA Langley Res. Center, Tech. Rep. NASA/TM-20230008547, 2025.

[25] N. Prins, "Easy, bias-free Bayesian hierarchical modeling of the psychometric function using the Palamedes toolbox," *Behav. Res. Methods*, vol. 56, pp. 485–499, 2024.

[26] M. Plummer, *JAGS: Just another gibbs sampler*, Astrophysics Source Code Library, ascl:1209.002, 2012.