SQAT: a MATLAB-based toolbox for quantitative sound quality analysis

Greco, G.F; Merino Martinez, R.; Osses, Alejandro; Langer, S.C

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the 52nd International Congress and Exposition on Noise Control Engineering

SQAT: a MATLAB-based toolbox for quantitative sound quality analysis

Gil Felix Greco[1]
Institut für Akustik, Technische Universität Braunschweig
Langer Kamp 19, 38106 Braunschweig, Germany

Roberto Merino-Martínez[2]
Faculty of Aerospace Engineering, Delft University of Technology
Kluyverweg 1, 2629 HS Delft, The Netherlands

Alejandro Osses[3]
Laboratoire des Systèmes Perceptifs, École Normale Supérieure, PSL University
29 Rue d'Ulm, 75005 Paris, France

Sabine C. Langer[4]
Institut für Akustik, Technische Universität Braunschweig
Langer Kamp 19, 38106 Braunschweig, Germany

ABSTRACT

Literature suggests that psychoacoustic metrics, which are established by models firmly based on the human auditory system, are able to better approximate perceptual attributes of sounds in a quantitative manner than conventional metrics based on the sound pressure level. Therefore, psychoacoustic metrics are commonly used as sound quality indicators in practical applications, such as product design or environmental noise assessment, in which human perception plays an important role. However, their calculation is not straightforward and the non-availability of verified, well-documented, open-source implementations in a single environment poses a challenge to reproducible research. In this contribution, we introduce the sound quality analysis toolbox (SQAT), an open-source compilation of MATLAB codes containing the implementation of the following key psychoacoustic metrics: loudness (ISO 532-1:2017), sharpness (DIN 45692:2009), roughness (Daniel & Weber's model), fluctuation strength (Osses et al. model), and tonality (Aures' model). A comprehensive description of their implementation is provided. Furthermore, a verification study is presented in order to demonstrate the reliability of the implementations, supporting the use of SQAT as a reliable tool for future works in the field of sound quality assessment based on psychoacoustic metrics.

[1] g.felix-greco@tu-braunschweig.de
[2] r.merinomartinez@tudelft.nl
[3] ale.a.osses@gmail.com
[4] s.langer@tu-braunschweig.de

1. INTRODUCTION

Most conventional sound metrics widely employed for sound design, assessment, and environmental law enforcement, are based on the physics-based sound pressure level (SPL or L_p). Several modifications to the SPL have been proposed in order to better represent the human hearing, such as the frequency-dependent human ear sensitivity by employing different weightings (for example, A-weighting and C-weighting) and the effect of sound duration by using the sound exposure level (SEL) [1]. Despite the improvements with respect to the simple but practical L_p metric achieved, literature suggests that more sophisticated sound quality (SQ) metrics are able to better approximate subjective ratings of loudness [2] and short-term annoyance [3] in a quantitative manner than conventional SPL-based metrics.

Ideally, listening experiments in controlled environments are required for evaluating the SQ of practical sounds [4, 5]. Nevertheless, listening experiments are costly and time-consuming and, thus, not suitable for exploring the typically large design spaces involved in the conceptual design phase. Alternatively, state-of-the-art SQ metrics established from psychoacoustic models are commonly used as quantitative SQ indicators in practical applications in which human perception plays an important role [6–8]. For example, psychoacoustic metrics can be combined into global perception-based metrics, such as the psychoacoustic annoyance [3, 9, 10], to quantitatively describe annoyance ratings obtained in listening experiments.

Despite the growing interest in the use of psychoacoustic metrics for quantitative SQ analysis, their calculation is not trivial and researchers in the field often experience a lack of publicly-available explanations, implementations, and validations of these auditory models. Therefore, the aim of this manuscript is to (1) introduce the sound quality analysis toolbox (SQAT), an open-source repository of MATLAB codes containing the implementation of key SQ metrics, and (2) provide a verification study of the main psychoacoustic metrics implemented on its first release.

2. SOUND QUALITY ANALYSIS TOOLBOX

SQAT [11] is an open-source repository of MATLAB codes containing the implementation of key metrics for quantitative SQ analysis. It should be noted that the implementations aim at reflecting the original models, but we cannot guarantee the exact same model outputs (and, thus, the estimated values) as when using the original codes. In the first release of SQAT (version 1.0) a collection of SQ metrics is provided under dedicated folders, as well as example (and associated sound files), validation, and utility codes. Aiming at facilitating the user's experience and allowing for expansion in future releases, the following folder structure is adopted (see Fig. 1):

– `psychoacoustic_metrics`: hosts a number of algorithms implementing specific psychoacoustic metrics, which are by default computed using calibrated input sound signals, in Pascal units. The algorithms are named using the following convention: `MetricName_FirstAuthorYear`;

– `sound_level_meter`: contains scripts to obtain the SPL vs. time using different frequency weightings (A, C, or Z) and time weightings (fast, slow, or impulse), and metrics derived from it. By default, the input is assumed to be a calibrated sound signal, in Pascal units;

– `examples`: in order to facilitate the initial use of the algorithms, an exemplary script is provided for each metric as a template;

– `sound_files`: contains WAV files required to run any codes within the `examples` folder;

– `utilities`: scripts that are complementary to any of the toolbox functions are provided here. The main utility scripts concern the calibration of the input signals and default input values for the implemented metrics;

– `validation`: contains scripts used to verify a particular algorithm; and

– `publications`: contains scripts to reproduce figures and/or tables of publications from the toolbox authors. All figures presented in this manuscript can be reproduced using the `pub_Greco2023_Internoise.m` code provided in this folder.

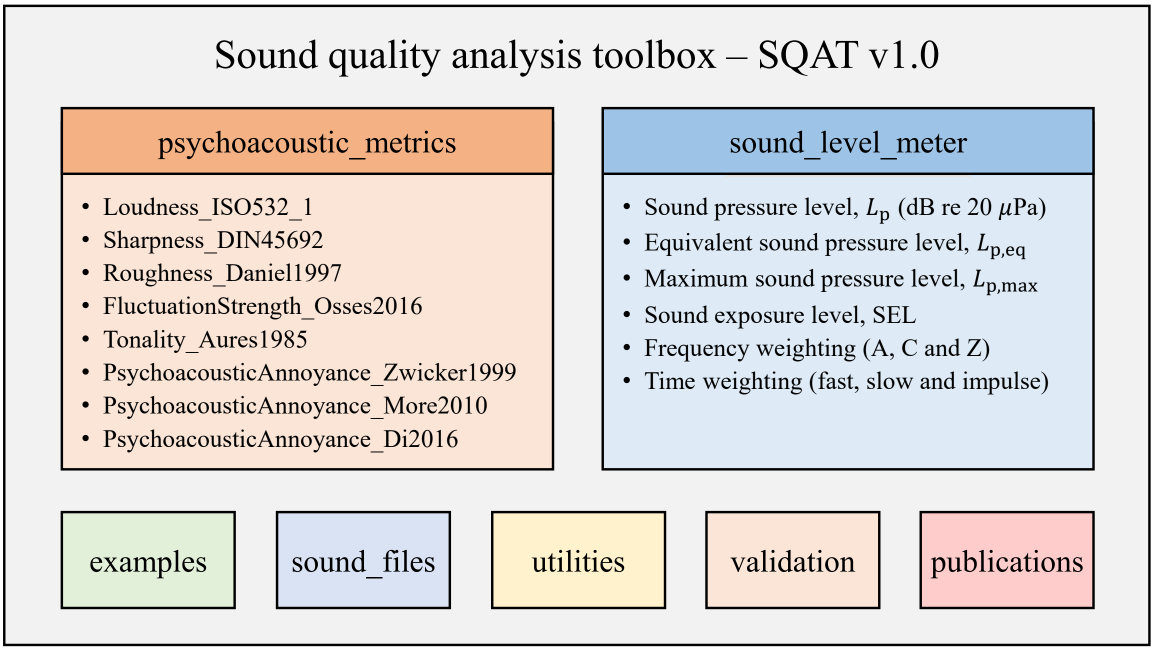


Figure 1. SQAT [11]: folder structure and metrics available in the first release.

3. PSYCHOACOUSTIC METRICS

In this section, a comprehensive description of the psychoacoustic metrics presented in Table 1 is provided. These are considered to be the main psychoacoustics metrics implemented in SQAT v1.0 whereas the three variants to estimate psychoacoustic annoyance [3, 9, 10], also presented in Fig. 1, are based on a combination of the main metrics.

Table 1. Overview of the psychoacoustic metrics described in Sec. 3 and associated models implemented in SQAT v1.0. The following parameters are used to describe the pure tone and amplitude-modulated (AM) reference signals: center frequency f_c, bandwidth BW, modulation frequency f_{mod}, and modulation depth m_d.

Metric	**Model**	**Reference signal**	**Value (unit)**
Loudness	ISO 532-1:2017 [12]	Pure tone, $f_c = 1$ kHz and $L_p = 40$ dB SPL	1 (sone)
Sharpness	DIN 45692:2009 [13]	Narrowband noise, $f_c = 1$ kHz, BW = 160 Hz and $L_p = 60$ dB SPL	1 (acum)
Roughness	Daniel & Weber [14]	AM tone, $f_c = 1$ kHz, $f_{mod} = 70$ Hz, $m_d = 1$ and $L_p = 60$ dB SPL	1 (asper)
Fluctuation strength	Osses *et al.* [15]	AM tone, $f_c = 1$ kHz, $f_{mod} = 4$ Hz, $m_d = 1$ and $L_p = 60$ dB SPL	1 (vacil)
Tonality	Aures [16]	Pure tone, $f_c = 1$ kHz and $L_p = 60$ dB SPL	1 (t.u.)

3.1. Loudness (ISO 532-1:2017)

Loudness, N, expressed in sone, quantifies the perceived magnitude of a sound in a linear scale (for example, a sound with 40 sone is perceived as twice as loud than a sound with 20 sone). In SQAT, the Zwicker method for stationary and time-varying sounds is implemented according to the ISO 532-1:2017 standard [12].

The loudness model described in Ref. [12] approximates the critical band rate z, which describes the frequency bandwidth of the human auditory "filter" in the Bark scale, using 1/3 octave-bands (OB). Therefore, in the first step, the input signal is decomposed into 28 1/3-OB, from 25 Hz to 12.5 kHz. Thereafter, corrections related to the critical band approximation by 1/3-OB are applied, leading to estimated critical band levels L_{CB} in 20 frequency bands instead of the 28 initial 1/3-OB.

In the following steps, the L_{CB} are adjusted by subtracting level corrections due to the ear's transmission characteristics a_0 and, in the case of a diffuse sound field, adding a level factor ΔL_{DF}. Additional level corrections are applied if the resulting L_{CB} individual values exceed the critical band level at the threshold in quiet L_{TQ} and the so-called core loudness N'_{C} is calculated using Eq. (1).

$$N'_{\mathrm{C}} = \left(0.0635 \cdot 10^{\,(0.025\,\cdot\,L_{\mathrm{TQ}})}\right) \cdot \left[\left(0.75 + 0.25 \cdot 10^{\,0.1\,\cdot\,(L_{\mathrm{CB}}-L_{\mathrm{TQ}})}\right)^{0.25} - 1\right] \tag{1}$$

Finally, the core loudness is mapped into 24 critical bands using a tabular and iterative procedure, where frequency masking effects are considered in order to obtain the unmasked specific loudness pattern $N'(z)$ (sone/Bark). The total loudness N, is obtained by integrating the specific loudness over the Bark scale, as given by Eq. (2). Once calculated, the total loudness can be converted into loudness levels, L_{N} (phon), in logarithmic scale using formulae provided by Ref. [12].

$$N = \int_0^{24\ \mathrm{Bark}} N'(z)\,\mathrm{d}z \quad \text{(sone)} \tag{2}$$

For time-varying sounds, three additional steps are considered: (i) the 1/3-OB over time are smoothed using third-order low-pass filters; (ii) the nonlinear temporal decay of the human hearing system is simulated after the calculation of the core loudness using an electrical circuit analogy; and (iii) the effects of temporal summation and forward masking on the total loudness are modeled using two first-order low-pass filters. In this case, the calculation provides a temporal sequence of specific loudness $N'(t,z)$ and total loudness $N(t)$ with a time resolution of 2 ms.

In SQAT, the implementation according to Ref. [12], named `Loudness_ISO532_1`, is an adaption of the code available in Ref. [17]. It computes the loudness from stationary or time-varying calibrated input signals, in Pascal units. An additional possibility is the use of stationary SPL in 1/3-OB as input. The exact method used for the loudness calculation needs to be specified by the user along with the sound field (free or diffuse).

3.2. Sharpness (DIN 45692:2009)

Sharpness, S, expressed in acum, is a measure of the high-frequency content of a sound. The greater the high-frequency content with respect to the rest of the spectral energy, the "sharper" or the "brighter" the sound is perceived. In SQAT, sharpness is implemented according to the DIN 45692:2009 norm [13], as

$$S = \mathrm{k} \cdot \frac{\int_0^{24\ \mathrm{Bark}} N'(z) \cdot g(z) \cdot z\,\mathrm{d}z}{\int_0^{24\ \mathrm{Bark}} N'(z)\,\mathrm{d}z} \quad \text{(acum)}\,, \tag{3}$$

where $0.105 \leq \mathrm{k} \leq 0.115$ is a constant to be adjusted so that the reference signal (see Table 1) yields a sharpness of one acum. In SQAT, $\mathrm{k} = 0.11$ is defined for all methods. Furthermore, $g(z)$ is a weighting function that emphasizes the specific loudness in the critical bands corresponding to high frequencies. Three weighting functions are defined by Ref. [13]. The standard one is given by Eq. (4).

In SQAT, the implementation according to Ref. [13] is named `Sharpness_DIN45692`. It computes the sharpness from calibrated input signals, in Pascal units, using by default the `Loudness_ISO532_1` code implemented in SQAT (see Sec. 3.1). Therefore, it requires inputs related to the loudness computation method (stationary or time-varying) and sound field (free or diffuse). For the sake of flexibility, an additional implementation named `Sharpness_DIN45692_from_loudness` is provided, only differing from the prior by using a previously calculated specific loudness (as a function of time or not) as input instead of a sound signal. In both implementations, one of the three weighting functions defined by the norm [13] needs to be specified by the user.

$$g(z) = \begin{cases} 0.15 \cdot e^{0.42(z-15.8)} + 0.85 & , z > 15.8 \text{ Bark} \\ 1 & , z \leq 15.8 \text{ Bark} \end{cases} \tag{4}$$

3.3. Roughness (Daniel & Weber's model)

Roughness, R, expressed in asper, is related to the hearing sensation caused by sounds with modulation frequencies between 15 Hz and 300 Hz, which are too rapid to be discerned separately [9]. Such rapid modulations give the impression of a "rough" or "harsh" sound, such as a lawnmower. Roughness has a band-pass characteristic as a function of f_{mod} and the carrier frequency, achieving its maximum at $f_{\text{mod}} = 70$ Hz for a 1 kHz pure tone [9]. Moreover, this metric is a function of the modulation depth and sound level [9].

In SQAT, an implementation of the roughness model from Daniel & Weber [14] is provided. This model is a modified version of the Aures model [18] to reduce the deviations between subjective and calculated values below the roughness just-noticeable difference[5] (JND) of 17 %. The Daniel & Weber's model can be summarized in three main stages. In the first stage, the transformation of the input signal into 47 excitation patterns in the basilar membrane is realized. In the second stage, each i^{th} excitation pattern is band-pass filtered in order to account for the roughness dependence on f_{mod} and a generalized modulation depth m_i^* is estimated. The specific roughness is calculated in the third stage as

$$R'_i = \mathrm{C_R} \cdot (g_\mathrm{R}(z_i) \cdot m_i^* \cdot k_{i-2} \cdot k_i)^2 \quad \text{(asper/Bark)}\,, \tag{5}$$

where $\mathrm{C_R} = 0.5$ is a calibration factor, $g_\mathrm{R}(z_i)$ is a weighting factor accounting for the roughness dependency on the carrier frequency and k is the cross-correlation coefficient between the band-pass filtered excitation patterns. Finally, the total roughness is obtained as

$$R = \mathrm{d}z \cdot \sum_{i=1}^{47} R'_i \quad \text{(asper)}\,, \tag{6}$$

where $\mathrm{d}z = 0.5$ Bark. Therefore, Eq. (6) expresses a relationship which is comparable to a discrete version of Eq. (2), but applied to obtain the total roughness as the area under the specific roughness.

The implementation of the Daniel & Weber's model [14] in SQAT, named `Roughness_Daniel1997`, is an adaption of the code available in Ref. [17]. It provides time-varying roughness $R(t)$ and specific roughness $R'(t, z)$ values with a resolution of 0.2 s from a calibrated input sound pressure signal, in Pascal units.

3.4. Fluctuation strength (Osses *et al.* model)

Fluctuation strength, FS, expressed in vacil, is associated with the hearing sensation caused by sounds with modulation frequencies up to approximately 20 Hz, achieving its maximum at $f_{\text{mod}} \approx 4$ Hz regardless the modulation type [9]. Because loudness changes slowly over time for such

[5]The roughness JND of 17 % is based on the work of Fastl & Zwicker [9], page 260, which states that an increment of roughness becomes audible for a 10 % increment of the modulation depth.

low modulation frequencies, the listener is able to discern each individual fluctuation, giving the impression of a "pulsating" or "beating" sound, such as an alarm siren. Similarly to roughness, FS has a band-pass characteristic as a function of f_{mod} and the carrier frequency, and it also depends on m_{d} and the sound level.

In SQAT, an implementation of the FS model from Osses *et al.* [15] is provided. This model is an adaption of the roughness model from Sec. 3.3. In comparison with the roughness model, the following modifications were employed by Ref. [15] to develop their model based on artificial stimuli with known FS values. In the first stage, the input signal is translated into excitation patterns using an analysis window of 2 s (90 % overlap) instead of 0.2 s in order to achieve a higher frequency resolution. Similarly to the roughness model, 47 excitation patterns corresponding to the outputs from the critical-band filter bank are obtained at the end of the first stage. In the second stage, a generalized modulation depth $m^{*}_{\mathrm{FS},i}$ is estimated for each i^{th} excitation pattern using a band-pass filter tailored to model the FS dependence on f_{mod}. The specific FS is obtained in the third stage as

$$\mathrm{FS}'_i = \mathrm{C}_{\mathrm{FS}} \cdot \left(g_{\mathrm{FS}}(z_i) \cdot (m^{*}_{\mathrm{FS},i})^{\mathrm{p_m}} \cdot |k_{\mathrm{FS},i-2} \cdot k_{\mathrm{FS},i}|^{\mathrm{p_k}} \right) \quad (\mathrm{vacil/Bark})\,, \tag{7}$$

where $\mathrm{C}_{\mathrm{FS}} = 0.498$, $\mathrm{p_m} = \mathrm{p_k} = 1.7$, k_{FS} is the normalized cross-variance coefficient between different auditory filters and $g_{\mathrm{FS}}(z_i)$ is a weighting factor that accounts for the carrier frequency. The total FS is obtained as the area under the specific FS using Eq. (8), where dz = 0.5 Bark.

$$\mathrm{FS} = \mathrm{dz} \cdot \sum_{i=1}^{47} \mathrm{FS}'_i \quad (\mathrm{vacil}) \tag{8}$$

In SQAT, the implementation of the FS model from Osses *et al.* [15] is denominated `FluctuationStrength_Osses2016`. It provides time-varying and specific FS values with a resolution of 0.2 s from a calibrated input sound pressure signal. Please note that, due to the length of the analysis window, which allows the model to appropriately extract modulation frequencies down to 1 Hz, it is recommended to analyze sounds with time length $t_{\mathrm{s}} \geq 2$ s. Nevertheless, for practical reasons, the implementation allows to analyze signals with smaller time lengths. In that case, the length of the analysis window is automatically set to the time length of the input signal. However, this limits the model to extract modulation frequencies down to $2/t_{\mathrm{s}}$ Hz.

3.5. Tonality (Aures' model)

Tonality is a term used to describe the degree to which a sound is perceived as tonal, which depends on many factors, such as bandwidth, center frequency, level above threshold, and the number of tonal components [19]. In SQAT, an implementation of the tonality metric developed by Aures [16] is provided.

Aures' tonality, K, expressed in tonal units (t.u.), quantifies how tonal a sound is perceived using a single value ranging from 0 to 1, which is computed based on a tonal weighting W_{T} and a loudness weighting W_{Gr} using Eq. (10). Initially, the spectrum of the signal is obtained and candidates for tonal components are identified based on the criteria established by Terhard *et al.* [20]. The estimated bandwidth, central frequency, and level of each i^{th} identified tone candidates are used to compute the tonal weighting as

$$W_{\mathrm{T}} = \sqrt{\sum_{i=1}^{N} [w'_1(\Delta z_i) \cdot w'_2(f_{\mathrm{c},i}) \cdot w'_3(\Delta L_i)]^2}, \tag{9}$$

where w_1, w_2, and w_3 are weighting functions accounting for the effect of the tones' bandwidth, in the Bark scale Δz, center frequency, and the sound pressure level excess,[6] ΔL, respectively. Finally, the

[6]The sound pressure level excess ΔL proposed by Terhard *et al.* [20] defines the perceptual relevance of tones based on masking effects and the hearing threshold at the central frequency of the tones.

weighting functions are obtained as $w'_n = w_n^{1/0.29}$ for $n = 1, 2, 3$.

The loudness weighting is calculated as $W_{\mathrm{Gr}} = 1 - (N_{\mathrm{Gr}}/N_{\mathrm{total}})$, where N_{Gr} is the total loudness of the input signal without the identified tonal components and N_{total} is the total loudness of the input signal. Finally, both weighting functions are combined and the Aures' tonality is calculated as

$$K = \mathrm{C} \cdot W_{\mathrm{T}}^{0.29} \cdot W_{\mathrm{Gr}}^{0.79} \quad (\mathrm{t.u.})\,, \tag{10}$$

where C is a calibration constant to be adjusted so that the reference signal (see Table 1) yields a tonality of one t.u. In SQAT, the calibration constant is defined as C =1.11.

The implementation of the Aures' tonality model [16] in SQAT is named `Tonality_Aures1985`. It provides time-varying tonality values with a resolution of 80 ms from a calibrated input sound pressure signal. It uses by default the `Loudness_ISO532_1` code implemented in SQAT (see Sec. 3.1). Therefore, the sound field (free or diffuse) used for the loudness calculation needs to be specified by the user.

4. VERIFICATION STUDY

During the development of SQAT, the implementation of each psychoacoustic metric described in Sec. 3 was verified based on the reference signals and corresponding unit values (see Table 1). The codes and signals used for this purpose are available, respectively in the `examples` and `sound_files` folders of the toolbox. Nevertheless, a verification based solely on the reference signals was observed to be insufficient to ensure the robustness and reliability of the codes. Moreover, standardized metrics require specific verification procedures. Therefore, a verification study using test sounds and reference values available in the literature is presented in this section. A more comprehensive verification of each algorithm is provided in the `validation` folder of the toolbox.

4.1. Loudness (ISO 532-1:2017)

The `Loudness_ISO532_1` implementation in SQAT is verified using the test signals[7] provided by the ISO standard [12]. The dataset contains 25 test signals with different characteristics, including synthetic signals (stationary and time-varying) and technical signals (time-varying).

Due to the extensive criteria and test signals required by the ISO standard [12] to verify the `Loudness_ISO532_1` implementation, the results presented hereafter comprehend only two selected cases (see Fig. 2). The complete verification can be found in the `validation` folder of the toolbox. For test signal 1 (stationary machinery noise, see Fig. 2a), the specific loudness results obtained using SQAT are in agreement with the tolerances stipulated by the ISO standard for the entire critical band rate. Moreover, the difference between the reference and calculated total loudness values is negligible. For test signal 14 (propeller-driven airplane, see Fig. 2b), which presents a complex temporal structure, the total loudness over time is computed by SQAT within the minimum tolerance requirements of ±5 % from the reference data along the entire time frame of the signal. For this case, the maximum total loudness N_{max} obtained using SQAT slightly overestimates the reference values by 0.24 sone or 1.05 % of the reference value.

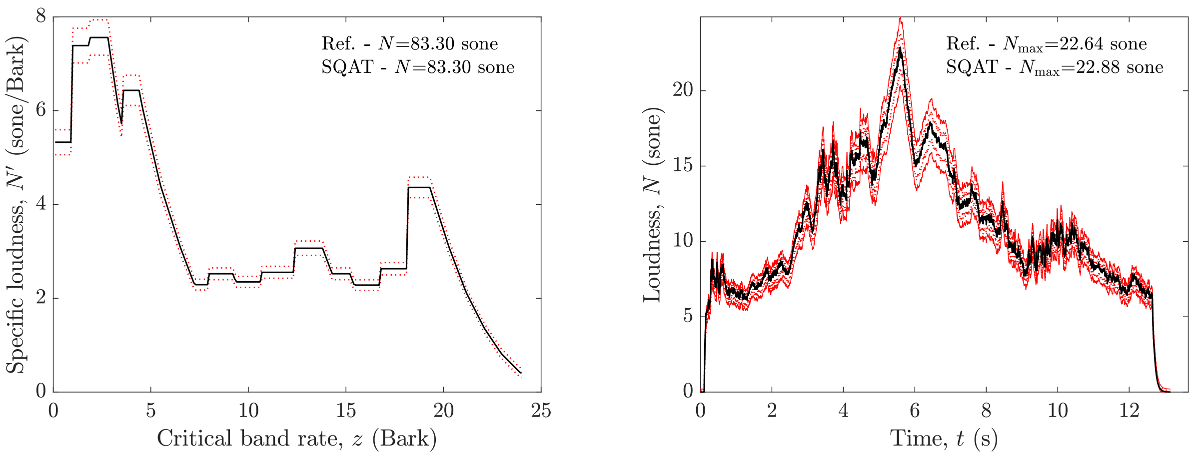
4.2. Sharpness (DIN 45692:2009)

The `Sharpness_DIN45692` implementation in SQAT is verified using the test signals provided by the DIN norm [13], which consists of: (i) 21 narrowband sounds (one-critical-band wide) with center frequencies in Bark (from 2.5 Bark to 22.5 Bark in 1 Bark steps), and (ii) 21 broadband sounds with variable lower frequencies in Bark (from 2.5 Bark to 22.5 Bark in 1 Bark steps) and fixed upper

[7]The dataset of test sounds provided by the ISO standard [12] is freely available and can be downloaded from the following link: http://standards.iso.org/iso/532/-1/ed-1/en.

frequency of 10 kHz ($\approx$ 22.4 Bark). All test signals are stationary noises with a total loudness of 4 sone.

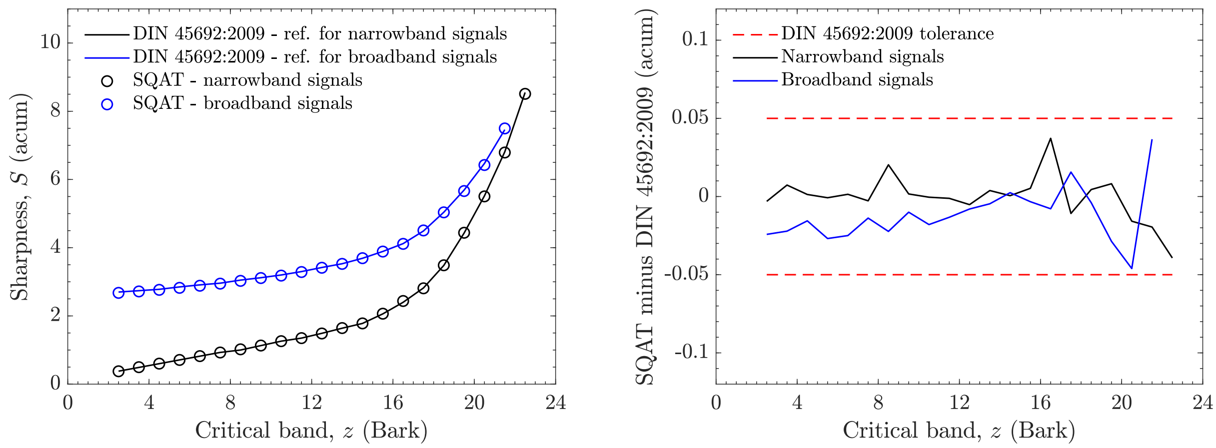
Figure 3 presents a comparison between the sharpness values obtained using the SQAT implementation and the reference values provided by the DIN norm [13] for each test signal, which are based on the standard weighting function defined by Eq. (4). In general, the SQAT implementation is able to provide adequate sharpness values for both narrowband and broadband test signals. The results presented in Fig. 3b indicate that the differences between the sharpness values provided by the `Sharpness_DIN45692` implementation in SQAT and the reference values are not bigger than 0.05 acum for any of the test signals, as required by the DIN norm [13].



(a) Test signal 1: stationary machinery noise.

(b) Test signal 14: propeller-driven airplane.

Figure 2. Verification of loudness implementation in SQAT according to ISO 532-1:2017 [12] for stationary and time-varying test sounds: $\pm 5\,\%$ tolerance (dashed red lines), and $\pm 10\,\%$ tolerance (solid red lines) reference values. Results obtained using SQAT are presented in black solid lines.



(a)

(b)

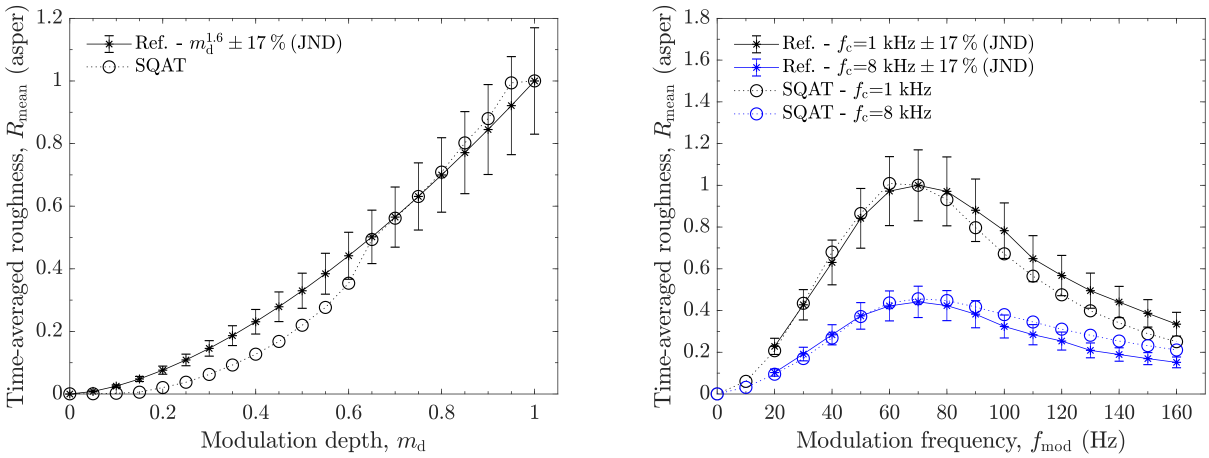
Figure 3. Verification of sharpness implementation in SQAT according to DIN 45692:2009 [13] for narrowband and broadband test signals: (a) absolute sharpness, and (b) difference between SQAT and reference values. The loudness of the test signals was obtained using the loudness method for stationary signals according to ISO 532-1:2017 [12], as implemented in SQAT.

4.3. Roughness (Daniel & Weber's model)

The `Roughness_Daniel1997` code implemented in SQAT is verified based on data available in the literature [9, 14] using AM signals, which were synthetically generated as

$$y(t) = \underbrace{[1 + m_\mathrm{d} \cdot \sin(2\pi f_\mathrm{mod} t)]}_{\text{Modulation waveform}} \cdot \underbrace{\sin(2\pi f_\mathrm{c} t)}_{\text{Carrier signal}}. \tag{11}$$

The sound pressure signal $p(t) = A \cdot y(t)$, in Pascal, is obtained by adjusting the level of the generated signal, $L_\mathrm{p,in} = 20\log_{10}(y_\mathrm{rms}(t)/20\,\mu\mathrm{Pa})$, based on a predefined $L_\mathrm{p,out}$, in dB SPL, using the calibration factor $A = 10^{(L_\mathrm{p,out} - L_\mathrm{p,in})/20}$. Figure 4 presents the results for AM signals with varying modulation depth and modulation frequency.



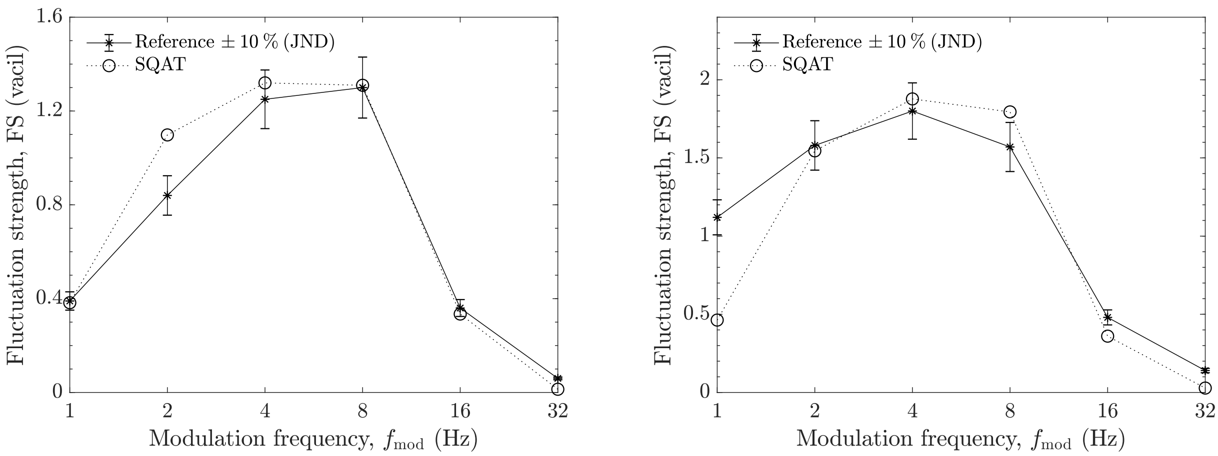
(a) AM tones ($f_\mathrm{c} = 1$ kHz, $f_\mathrm{mod} = 70$ Hz and $L_\mathrm{p} = 60$ dB SPL). (b) AM tones ($m_\mathrm{d} = 1$ and $L_\mathrm{p} = 60$ dB SPL).

Figure 4. Roughness model from Daniel & Weber [14]: verification of the implementation in SQAT for (a) roughness as a function of modulation depth compared with the power law suggested by Fastl & Zwicker [9], and (b) roughness as a function of modulation frequency compared with reference data from listening tests published in Ref. [14]. The error bars express the roughness JND also reported in Ref. [9].

The results presented in Fig. 4a show that the implementation in SQAT provides roughness values following the reference power law for $0.7 \leq m_\mathrm{d} \leq 1$. Below this range, the implementation provides underestimated values outside of the roughness JND. Figure 4b shows that the `Roughness_Daniel1997` implementation is able to adequately model the band-pass characteristic of roughness as a function of the modulation frequency for distinct frequency carriers. The maximum roughness value achieved at $f_\mathrm{mod} = 70$ Hz for a 1-kHz AM tone is correctly predicted by SQAT. Moreover, with the exception of few cases within $130\,\mathrm{Hz} \leq f_\mathrm{mod} \leq 160\,\mathrm{Hz}$, the differences between the reference data and the values obtained using SQAT are smaller than the roughness JND. A complete validation comprising the roughness dependence on the modulation frequency is provided in the `validation` folder of the toolbox for AM tones with different carrier frequencies, as well as frequency modulated (FM) tones, and white noise.

4.4. Fluctuation strength (Osses *et al.* model)

Artificial AM tones and AM broadband test signals were used for the verification of the `FluctuationStrength_Osses2016` code implemented in SQAT. These sound stimuli, which were used in Ref. [15] to develop the original FS model together with FM tones, are available in Ref. [21]. The results presented in Fig. 5 indicate that the model implementation is able to provide reasonable FS estimates for both types of test signals under consideration. This is especially the case for AM tones with $4\,\text{Hz} \leq f_{\text{mod}} \leq 32\,\text{Hz}$ (see Fig. 5a) and AM broadband noises with $2\,\text{Hz} \leq f_{\text{mod}} \leq 4\,\text{Hz}$ (see Fig. 5b). The observed differences between the reference values and the model estimates seem to be tolerable, as the estimates follow the general FS trends. Additional verification results of the `FluctuationStrength_Osses2016` implementation are provided in Ref. [15] for FM tones and everyday sounds.



(a) AM tones ($f_c = 1$ kHz, $m_d = 1$ and $L_p = 70$ dB SPL).

(b) AM broadband noises ($f_c = 1$ kHz, BW = 16 kHz, $m_d = 1$ and $L_p = 60$ dB SPL).

Figure 5. Fluctuation strength model from Osses *et al.* [15]: verification of the implementation in SQAT for AM sounds as a function of modulation frequency. The fluctuation strength values from Fastl & Zwicker [9] are adopted as reference. The error bars express the fluctuation strength JND also reported in Ref. [9].

4.5. Tonality (Aures' model)

Despite the thorough use of the Aures' tonality metric in the literature [3, 19], not many reproducible cases are available to verify its implementation. Therefore, the verification presented in Fig. 6 is limited to the reference signal (see Table 1), and pure tones with different signal-to-noise ratios (SNR) from narrowband noises. The results presented in Fig. 6a show that the `Tonality_Aures1985` implementation is able to adequately compute the value of 1 t.u. for the reference signal. Figure 6b shows that the implementation in SQAT is able to provide tonality results with an error smaller than 0.1 t.u. for the entire range assessed. For sounds with $0\,\text{dB} \leq \text{SNR} \leq 30\,\text{dB}$, the algorithm has difficulties finding the tones consistently along all time-frames, which compromises the estimation of the time-averaged tonality value. Modifications on the frequency resolution and on the criteria used to find the tones, as suggested by Estreder *et al.* [22], did not led to any improvements. For very high SNRs (above 40 dB), the absolute differences between the calculated and reference results are not bigger than 0.018 t.u.

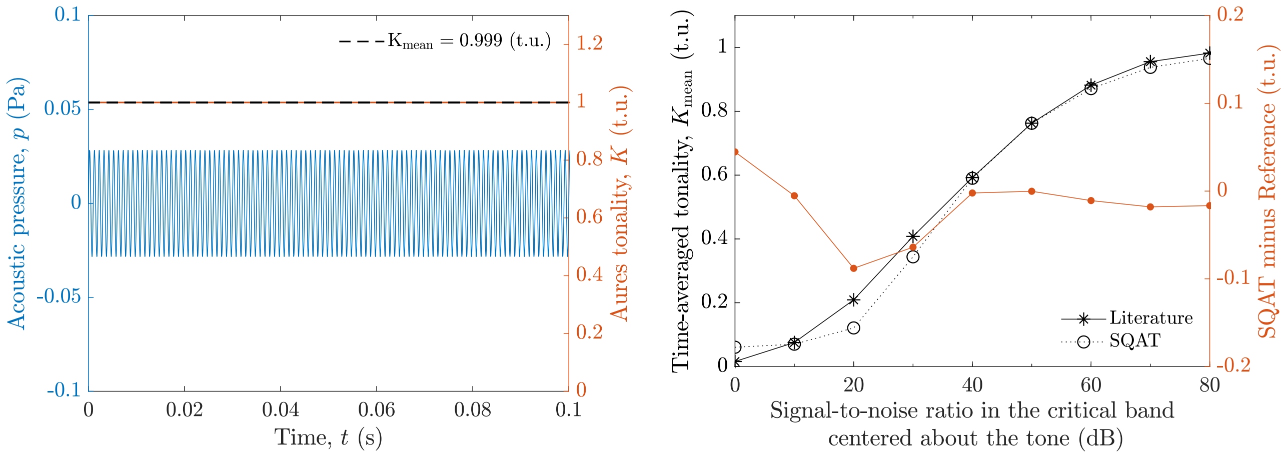
5. CONCLUSIONS

This manuscript introduced SQAT [11]: an open-source sound quality analysis toolbox for MATLAB. The toolbox structure was described as well as the implemented metrics available in the first release. Furthermore, a verification study of the main psychoacoustic metrics implemented in the toolbox was presented.

The results of the verification study presented in Sec. 4 provide evidence on the reliability and limitations of the implementations in a transparent manner, supporting their use for quantitative

SQ assessment on sound signals. All verification analyses presented in this manuscript can be reproduced using the `pub_Greco2023_Internoise.m` code provided in the `publications` folder of the toolbox. Furthermore, additional verification studies can be found in the `validation` folder.

With the release of this open-source toolbox, we hope to contribute to the scientific community by providing a readily accessible and reliable tool that allows for a quick yet representative quantitative SQ analysis. In addition, we expect to favor reproducible science as the number of studies using SQAT grows. In future releases, additional features and metrics, such as the effective perceived noise level (EPNL) and speech transmission index (STI), are foreseen to be included. Constructive feedback on how to improve SQAT is certainly appreciated and users willing to collaborate by including additional SQ metrics are warmly welcome.



(a) Pure tone ($f_c = 1$ kHz and $L_p = 60$ dB SPL).

(b) Pure tone ($f_c = 1$ kHz and $L_p = 85$ dB SPL) with different prominence SPLs from an narrowband noise (one-critical-band wide, BW = 1081–918.8 Hz).

Figure 6. Tonality model from Aures [16]: verification of the implementation in SQAT for (a) reference signal, and (b) pure tones with different SNRs from narrowband noises compared with the reference values from Hastings *et al.* [19].

ACKNOWLEDGEMENTS

The authors would like to acknowledge the funding provided by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2163/1 – Sustainable and Energy Efficient Aviation – Project-ID 390881007 and by the Agence Nationale de la Recherche (ANR, French National Research Agency) – Grant ANR-17-EURE-0017.

REFERENCES

[1] D. S. Houser, W. Yost, R. Burkard, J. J. Finneran, C. Reichmuth, and J. Mulsow. A review of the history, development and application of auditory weighting functions in humans and marine mammals. *The Journal of the Acoustical Society of America*, 141(3):1371–1413, 2017.

[2] H. Fastl and U. Widmann. Subjective and physical evaluation of aircraft noise. *Noise Control Engineering Journal*, 35(2):61–63, 1990.

[3] S. R. More. *Aircraft noise characteristics and metrics*. PhD thesis, Purdue University, 2010.

[4] R. Merino-Martínez, R. Pieren, and B. Schäffer. Holistic approach to wind turbine noise: From blade trailing-edge modifications to annoyance estimation. *Renewable and Sustainable Energy Reviews*, 148, 2021.

[5] R. Pieren, L. Bertsch, D. Lauper, and B. Schäffer. Improving future low-noise aircraft technologies using experimental perception-based evaluation of synthetic flyovers. *Science of The Total Environment*, 692:68–81, 2019.

[6] R. Merino-Martinez, A. Vieira, M. Snellen, and D. G. Simons. Sound quality metrics applied to aircraft components under operational conditions using a microphone array. In *25^{th} AIAA/CEAS Aeroacoustics Conference, Delft, The Netherlands*, 2019.

[7] G. Felix Greco, L. Bertsch, T. P. Ring, and S. C. Langer. Sound quality assessment of a medium-range aircraft with enhanced fan-noise shielding design. *CEAS Aeronautical Journal*, 12(3):481–493, 2021.

[8] M. A. Boucher, S. Krishnamurthy, A. W. Christian, and S. A. Rizzi. Sound quality metric indicators of rotorcraft noise annoyance using multilevel analysis. *The Journal of the Acoustical Society of America*, 153(2):867–876, 2023.

[9] H. Fastl and E. Zwicker. *Psychoacoustics: facts and models, Third edition*. Springer-Verlag, Berlin, 2007.

[10] G. Q. Di, X. W. Chen, K. Song, B. Zhou, and C. M. Pei. Improvement of zwicker's psychoacoustic annoyance model aiming at tonal noises. *Applied Acoustics*, 105:164–170, 2016.

[11] G. Felix Greco, R. Merino-Martínez, and A. Osses. SQAT: a sound quality analysis toolbox for MATLAB, Zenodo, 2023. https://doi.org/10.5281/zenodo.7934709.

[12] ISO 532-1: Acoustics - Methods for calculating loudness - Part 1: Zwicker method. Standard, International Organization for Standardization, ISO, Geneva, Switzerland, 2017.

[13] DIN 45692: Measurement technique for the simulation of the auditory sensation of sharpness. Standard, Deutsches Institut für Normung: Berlin, Germany, Beuth, Verlag, 2009.

[14] P. Daniel and R. Weber. Psychoacoustical roughness: Implementation of an optimized model. *Acta Acustica united with Acustica*, 83(1):113–123, 1997.

[15] A. Osses Vecchi, R. García León, and A. Kohlrausch. Modelling the sensation of fluctuation strength. *Proceedings of Meetings on Acoustics*, 28(1), 2016.

[16] W. Aures. Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale (A model for calculating the sensory euphony of various sounds). *Acta Acustica united with Acustica*, 59(2):130–141, 1985.

[17] Cabrera, D. and Jimenez, D. and Martens, W. L. Audio and Acoustical Response Analysis Environment (AARAE): A tool to support education and research in acoustics. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 249(6):1667–1676, 2014.

[18] W. Aures. Ein berechnungsverfahren der rauhigkeit (A calculation method of roughness). *Acta Acustica united with Acustica*, 58(5):268–281, 1985.

[19] A. Hastings, K. H. Lee, P. Davies, and A. M. Surprenant. Measurement of the attributes of complex tonal components commonly found in product sound. *Noise Control Engineering Journal*, 51(4):195–209, 2003.

[20] E. Terhardt, G. Stoll, and M. Seewann. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *The Journal of the Acoustical Society of America*, 71(3):679–688, 1982.

[21] G. Felix Greco, R. Merino-Martínez, and A. Osses. SQAT v1_0: dataset of validation sounds, Zenodo, 2023. https://doi.org/10.5281/zenodo.7933206.

[22] J. Estreder, G. Piñero, M. de Diego, J. Rämö, and V. Välimäki. Improved aures tonality metric for complex sounds. *Applied Acoustics*, 204:109238, 2023.