DOI: 10.61782/fa.2025.0951

FORUM ACUSTICUM
EURONOISE 2025

VERIFIED IMPLEMENTATIONS OF THE SOTTEK PSYCHOACOUSTIC HEARING MODEL STANDARDISED SOUND QUALITY METRICS (ECMA-418-2 LOUDNESS, TONALITY AND ROUGHNESS)

Michael J B Lotinga[1*] **Matt Torjussen**[2] **Gil Felix Greco**[3]

[1] Acoustics Research Centre, University of Salford, UK
[2] ANV Measurement Systems, Milton Keynes, UK
[3] Institut für Akustik und Dynamik, Technische Universität Braunschweig, Germany

ABSTRACT

Metrics for describing perceptual sound qualities have previously been developed from a variety of psychoacoustic models. The Sottek Hearing Model (SHM) offers a holistic approach to characterising perceptual sound quality and the SHM metrics for loudness, tonality and roughness have been defined in an international standard, ECMA-418-2. Implementations of the ECMA-418-2 standardised sound quality metrics have been developed, documented and verified using both simple test signals and field audio recordings of complex sound scenes, and have also been made available in the open-source software package 'SQAT' (Sound Quality Analysis Toolbox). The metrics, implementation, application and usage guidance are outlined alongside presentation of example verification cases, demonstrating the accuracy and capabilities of the tools.

Keywords: *psychoacoustic model, sound quality metric, loudness, tonality, roughness*

1. INTRODUCTION

To explore and understand auditory perception, psychoacoustic models are essential, from which metrics to quantify sound quality (SQ) attributes, such as loudness, sharpness, roughness, tonality and fluctuation strength are established. Over time, various SQ metrics (SQMs) have been developed and refined from a range of perceptual models. Most commonly, each model development has focussed on one specific attribute in a piecemeal fashion, leading to the common practice today of analysing psychoacoustics using a 'patchwork' of metrics with varying assumptions and approaches. The Sottek Hearing Model (SHM) represents an attempt to model auditory perception in a holistic manner, which uses a common basis as a platform for developing metrics describing a range of SQs. The core of the model is the estimation of a 'basis' loudness, computed as a non-linear transformation of the bank of excitation signals corresponding with basilar membrane frequency selectivity, following outer and middle ear filtering. A detailed description of the theoretical basis and empirical support for the SHM has been presented by Sottek [1–4].

The SHM-based metrics for tonality T, loudness N and roughness R are standardised in the 3rd edition of ECMA-418-2 [5]. As described in the following, the SHM T, N and R metrics have been implemented in open-source MATLAB codebases, and verified using both simple artificial signals and outdoor recordings of complex sound scenes. The implementation was developed during research undertaken as part of the RefMap project [1] and maintained in an open-access repository [6]. This implementation has subsequently been amalgamated into the open-source sound quality analysis toolbox 'SQAT' v1.3 [7], which also offers verified implementations of other SQMs [8]. The tools described here are be-

*__Corresponding author__: m.j.lotinga@edu.salford.ac.uk.

[1] https://www.refmap.eu

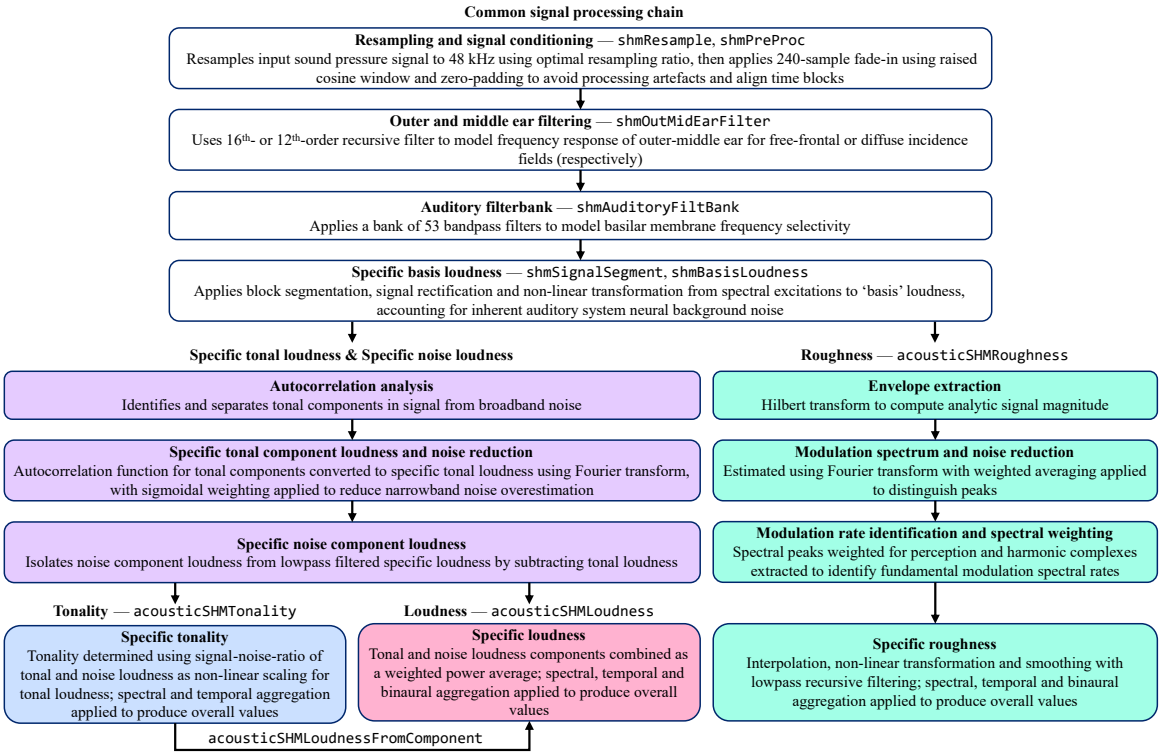


Figure 1: Overview of the implementated ECMA 418-2 algorithms.

lieved to represent the first open-source implementation of the ECMA-418-2 SQMs that has been fully verified using both simple and complex sounds.

2. METRIC IMPLEMENTATIONS

The implementation of the SHM SQMs follows ECMA-418-2:2024 [5], which sets out the signal processing chain for each metric. As illustrated in Fig. 1, the three calculated metrics share a common series of stages yielding the specific basis loudness N'_{basis}, which is the time-dependent loudness in each of the critical bands used in the SHM (denoted z). From the N'_{basis}, the metric-specific processing begins. The implemented tools automatically handle single- or two-channel signals. For the latter, the N and R algorithms offer the option for an additional binaurally-combined output as well as the channel-specific results.

2.1 Common signal processing chain

The common signal processing comprises four key stages. The `shmResample` function ensures that all the processing functions and filters operate on a consistent basis at the required standard rate of 48 kHz. The signal is further conditioned by the `shmPreProc` function, which applies fade-in and zero-padding, reducing processing artefacts due to signal discontinuities and filter behaviours. The zero-padding also maintains synchronisation in processing over time blocks segmented later in the process.

The frequency-dependent attenuation of the outer and middle ear is implemented in the `shmOutMidEarFilter` function using a recursive filter. The filter applied depends on the user-selected soundfield: free-field frontal incidence is modelled using 16$^{\text{th}}$-order, implemented as 8 cascaded 2$^{\text{nd}}$-order filters, while diffuse field incidence is modelled by omitting the first 3 2$^{\text{nd}}$-order stages. A note in ECMA-418-2 highlights that the inclusion of the outer-middle ear filtering stage implies that binaural recordings are assumed to be equalised to compensate for the effect of the head and ears.

The `shmAuditoryFiltBank` function creates a bank of 53 bandpass filters modelling the basilar membrane frequency selectivity. The implementation employs recursive

difference equations for each band. Each critical band filter overlaps by half a critical band, which is mapped to frequency based on a function approximating the Zwicker definition [9].

The final stage in the common signal processing chain is the transformation of the bandpass-filtered excitation signals to N'_{basis}. First, the signals in each band are segmented into time blocks. As well as facilitating parallel processing of blocks and overlapping to reduce the influence of noise during processing, this also enables each metric to divide the signal into suitably-sized blocks according to the resolution required in the spectral domain: the R metric uses longer time blocks, while the T and N metrics require a variable block length that depends on the spectral band (with lower bands requiring longer block lengths). The segmented signals are then rectified and transformed with a function that represents compressive auditory non-linearity, which leads to a relationship between sound pressure and SHM N that has been demonstrated to agree more closely with current equal-loudness contours than the existing ISO-standardised N models [10]. The common output at this juncture is N'_{basis}, used as input to the T, N and R metric-specific paths.

2.2 Tonality

As shown in Fig. 1, the signal processing path for the T and N metrics continues to be shared for three further key stages. To avoid duplicating code over separate functions, this is addressed by the automatic calling of the T metric function `acousticSHMTonality` from the N metric function (`acousticSHMLoudness`).

Autocorrelation analysis forms the core of the calculation, which identifies tonal components within stochastic noise by exploiting the periodicity of the autocorrelation function (ACF) for periodic signal components. Averaging across adjacent critical bands is performed to reduce noise in the ACF spectra.

Following this averaging, an initial estimation of specific tonal loudness is obtained by identifying the largest tones and corresponding frequencies within the ACF spectra. A series of noise reduction steps are then applied to refine the estimation of the loudness of tonal content, using a low-pass filter (`shmNoiseRedLowPass`) with a sigmoidal weighting function. This yields the final estimate of the specific loudness of tonal components.

The specific noise loudness is extracted by subtracting the specific tonal loudness from a lowpass-filtered representation of the total specific loudness obtained from the ACF. These two components, the specific noise and tonal loudnesses, can then be passed to the loudness processing to obtain the full psychoacoustic loudness, described in Section 2.3.

To derive the tonality, the component specific loudnesses are used to calculate a signal-noise-ratio (SNR) for each time block, based on the most prominent tonal component relative to the spectral sum of the noise component. This SNR is applied in a non-linear scaling function that yields the time-dependent specific tonality $T'(t,z)$. The $T'(t,z)$ then forms the basis for aggregating over critical bands and time to obtain the overall tonality T, time-dependent tonality $T(t)$ or time-averaged specific tonality $T'(z)$. The time aggregation is performed as a weighted arithmetic mean value, with values $\leq$0.02 tu_{HMS} excluded from the averaging.

2.3 Loudness

Since most of the processing to obtain N takes place in the path shared with T, there are very few steps required to produce the full N output. The specific component tonal and noise loudness are combined using a weighted power average that also depends on N'_{basis}. This yields the full specific loudness $N'(t,z)$, which is then aggregated to produce N, $N(t)$ and $N'(z)$. The aggregation over time is based on another power average approach, which has shown better agreement with perceptual results than the '5%-exceeded' (95$^{\text{th}}$ percentile) aggregation that has historically been commonly used for the Zwicker loudness model [11, 12].

Since the loudness function `acousticSHMLoudness` uses almost entirely the same processing as the `acousticSHMTonality` function, an efficiency shortcut tool is also provided to reduce processing time when calculating both tonality and loudness for the same signal. This function, `acousticSHMLoudnessFromComponent` (Fig. 1), takes as input the component tonal and noise loudness outputs from `acousticSHMTonality`, thus avoiding the redundancy of running the same signal processes twice. This tool effectively halves the processing time when calculating both metrics, since the remaining operations following calculation of the component loudnesses are trivial, in terms of computational cost.

2.4 Roughness

The roughness metric is implemented in the `acousticSHMRoughness` function, which begins with Hilbert transform envelope extraction from the auditory filtered excitation signals. This step is followed by downsampling to reduce the sample rate to 1.5 kHz.

This preserves the critical modulation phase relationships while optimising computational efficiency. The envelope spectra in each band are obtained via a windowed Fourier transform that is scaled using the N'_{basis}; this yields a set of bandpass loudness-normalised modulation spectra. To suppress extraneous noise in these spectra, band averaging is performed, followed by a non-linear weighting that emphasises the modulation peaks.

The next stage involves identifying modulation rates and weighting the modulation spectral peaks using perceptual weighting functions. Modulation rates are identified by locating prominent local maxima in the modulation spectra, the estimates from which are then refined and weighted for perception of high modulation rates. Then, the perception-weighted modulation spectra are analysed to identify harmonically related modulation rates, and these 'harmonic complexes' highlight the fundamental modulation rates. The fundamental rates are used to apply another perceptual weighting for low modulation rates. Both weightings (for low or high modulation rates) are implemented in the `shmRoughWeight` function.

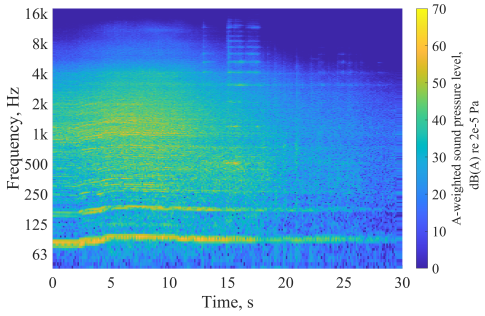
An optional 'entropy weighting' stage is defined in ECMA-418-2, which can be useful for applications involving rotating sources, and requires input of a measured rotational speed time series that corresponds with the audio signal. At present, this option is not included in the implementation.

Following the rate extraction and weighting stage, the weighted modulation amplitudes are resampled down to the output rate of 50 Hz, resulting in an initial estimate of the specific roughness. A non-linear transformation is applied to account for different signal bandwidths, the result of which is calibrated. The output is further refined via an asymmetric smoothing with recursive filtering operations implemented in `shmRoughLowPass`, yielding the time-dependent specific roughness $R'(t,z)$. This is then aggregated to produce R, $R(t)$ and $R'(z)$, wherein R is derived as the '10%-exceeded' value (90$^{\text{th}}$ percentile), and $R'(z)$ is the arithmetic average over time.

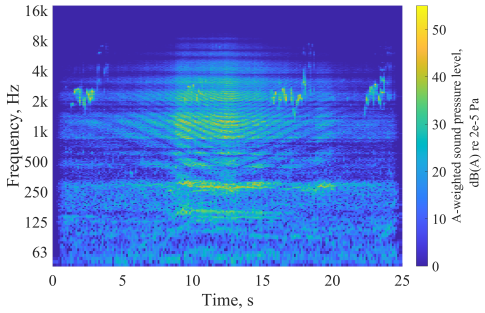
3. VERIFICATION

ECMA-418-2:2024 [5] does not provide test signals or criteria to verify a specific implementation. Therefore, the verification presented here is performed by comparison with the results from commercial software (Artemis Suite 15.7 from HEAD acoustics), which are adopted as the reference.

The verification examples are obtained using excerpts from two recorded outdoor scenes, spectrograms for which are shown in Fig. 2: (signal 2a) a railway station featuring a train departing, and (signal 2b), an urban park being used for leisure, within which the sound of an unmanned aircraft system (UAS, or 'drone') flying overhead has been added using auralisation processing. Both of the original recordings were obtained during the EigenScape project [13], while the UAS auralisation was part of work undertaken during the RefMap project [14]. For brevity, only a selection of the results are presented; further examples can be obtained by running the MATLAB script used to generate the verification figures [2]. Additional verification studies using artificial and recorded signals are also available in the code repositories of RefMap [3] and SQAT[2].



(a)



(b)

Figure 2: Spectrograms of audio signals used for presented verification: (a) train station, and (b) park with UAS overflight.

Analysis of the train station recording (Fig. 2a)

[2] https://github.com/ggrecow/SQAT

[3] https://github.com/acoustics-code-salford/refmap-psychoacoustics.

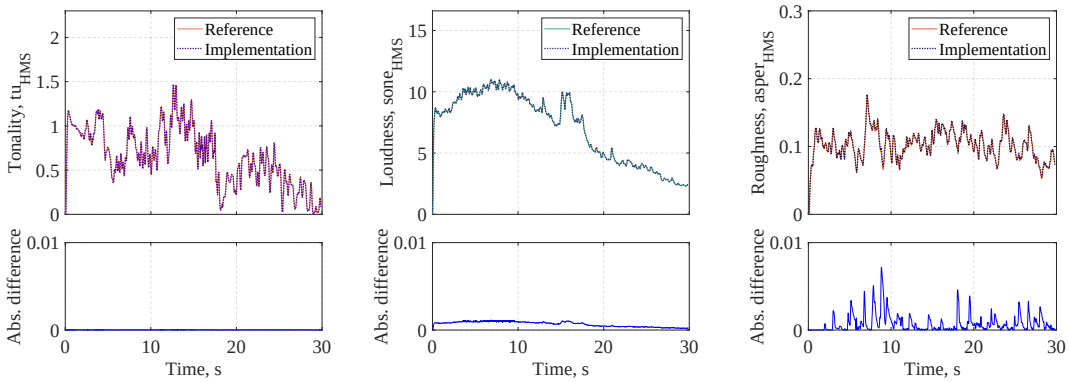


Figure 3: Comparison of results for sound 2a from commercial software (reference) and the implemented algorithms for different time-dependent sound quality metrics: (a) tonality, (b) loudness, and (c) roughness.

shows strong broadband sound energy over the 250–4000 Hz range, accompanied by two non-stationary harmonic tones emerging in the low frequency range; these features coincide with a train departing the platform. At 13–18 seconds into the signal, further harmonic tones are visible in the high-frequency range 3–14 kHz, which also exhibit transient impulses at their onset — these features coincide with rail squeal sounds generated by interactions with the train wheels.

The park recording (Fig. 2b) exhibits quasi-stationary low-level broadband sound, punctuated by prominent transient events in the 1.5–3.5 kHz range, with a rising pattern up to ∼6 kHz. The transient features are bird calls in the foreground, while the steady broadband energy comprises contributions from faint road traffic and indistinct leisure activity. The UAS overflight can be seen to emerge in the middle of the recording, with the Doppler shift and ground effect interference bands that are characteristic of aviation sources.

The results for time-dependent overall SQM values corresponding with sound 2a are presented in Fig. 3. The implemented algorithms provide results in good agreement with the reference in all cases, adequately capturing fast magnitude changes over time caused by the dynamic nature of the sound signal. This demonstrates the ability of the algorithms to provide suitable outcomes for technical sounds, which have far more complex temporal structures than most stationary sounds.

The accuracy of the time-dependent results presented in Fig. 3 are quantitatively measured based on the absolute difference from the reference results. It is shown in Fig. 3 that absolute differences in time-dependent tonality, loudness and roughness between the reference and the present implementations are all below 0.01. There are no implementation tolerances defined within the current version of ECMA-418-2. For implementations of Zwicker loudness, ISO 532-1:2017 [12] defines a verification criterion of 0.1 sone for time-dependent loudness and 0.1 sone/Bark for specific loudness from the published reference data. For context, a perceptual just-noticeable-difference (JND) for subjective loudness of stationary white noise is around 0.5 dB for sound pressure levels $\geq$ 40 dB [9], which is equivalent to ∼0.1 sone$_{HMS}$. A JND for subjective roughness is thought to be around 0.17 asper [9], while there is no published JND for tonality known to the authors.

The human auditory system is not equally sensitive to all frequencies within the range of audible sound. This frequency-dependent characteristic of the human ear becomes an important factor when modelling different perceptual dimensions of sound. For example, tonality perception depends on the centre frequency and bandwidth of prominent tonal components, peaking at around 1.5 kHz for pure tones [9] while, for a 1-kHz amplitude modulated tone, the maximum roughness sensation occurs at modulation rates of 70 Hz [9]. Therefore, SQ analysis in the frequency domain is crucial in many practical applications, as it helps determining aurally relevant spectral components which are otherwise not unveiled by time-dependent

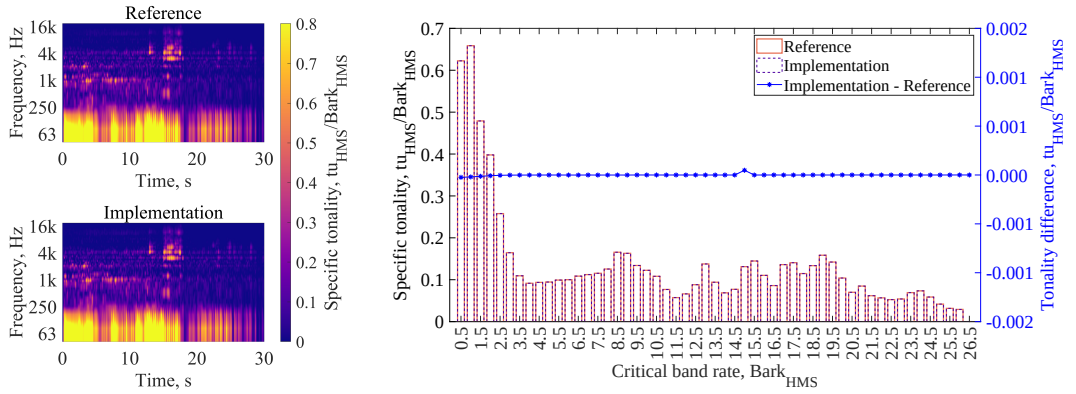


Figure 4: Comparison of (a) time-dependent and (b) time-averaged specific tonality results for sound 2a from commercial software (reference) and the implemented algorithms.

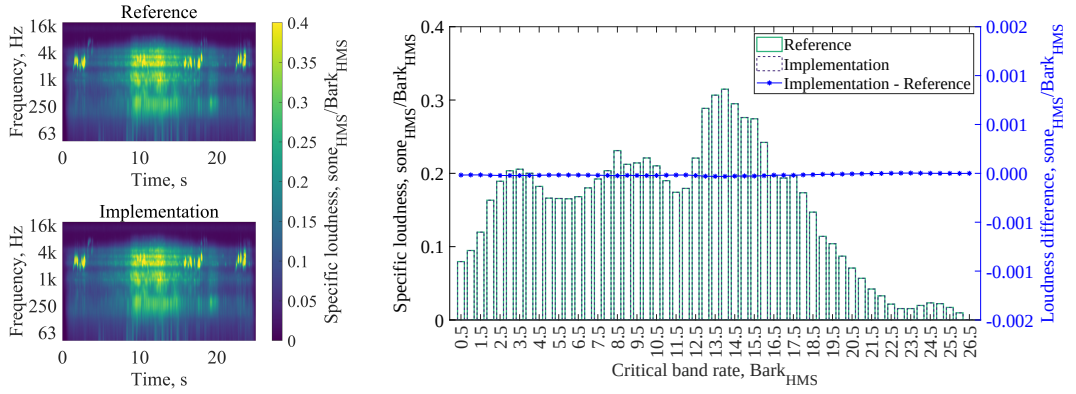
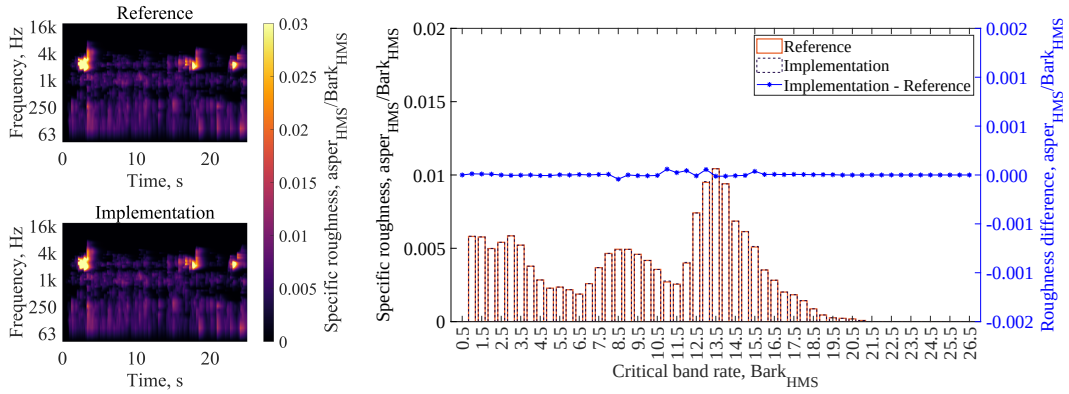


Figure 5: Comparison of (a) time-dependent and (b) time-averaged specific loudness results for sound 2b from commercial software (reference) and the implemented algorithms.

quantities. This is exemplified in Fig. 4, where results for specific tonality are presented. Based on the $T'(t,z)$ results shown in Fig. 4a, it is possible to obtain a full picture about the tonality of sound 2a from the temporal and spectral characteristics. The $T'(t,z)$ complements the $T(t)$ results from Fig. 3a not only by helping to identify the most relevant spectral components contributing to tonality, but also when they occur in time. For example, it is possible to observe in Fig. 4a that the low-frequency tonality due to the train-motor sound is prominent during almost the entire duration of the sound signal, while the rail squeal tonality at ∼3–5 kHz is also clearly highlighted by the metric at ∼15 seconds. Fig. 4a qualitatively demonstrates that the results obtained using the implemented tonality algorithm are in good agreement with the reference results from the commercial software.

Information about tonality in the frequency domain can also be analysed in terms of $T'(z)$, which is presented in Fig. 4b. This metric is useful when analysing dynamic signals recorded for a long period of time, as it indicates in which bands tonality occurred the most along the entire sound duration. However, information about any promi-



(a) (b)

Figure 6: Comparison of (a) time-dependent and (b) time-aggregated specific roughness results for sound 2b from commercial software (reference) and the implemented algorithms.

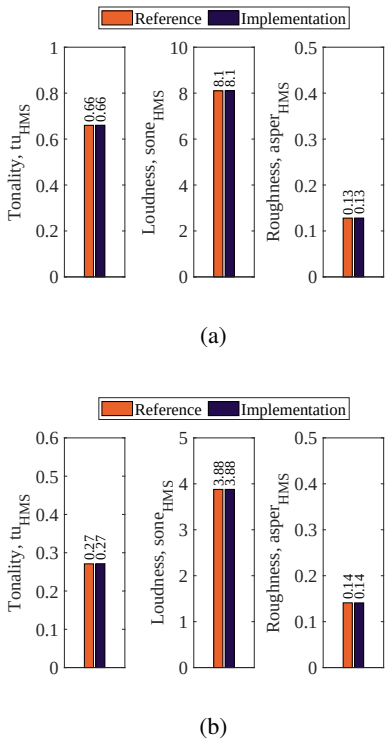


Figure 7: Comparison of overall values obtained for different sound quality metrics: (a) sound 2a, and (b) sound 2b.

nent tonal components occurring for short-time periods is obscured. Regarding the accuracy of the implemented algorithm, the results are in very close agreement with the reference.

Results for specific loudness analysis of sound 2b are shown in Fig. 5. In Fig. 5a, the transient bird call sounds are shown to be approximately equally loud within the corresponding frequency range (∼3 kHz) as the peak of the UAS overflight. The aggregated values shown in Fig. 5b also show the greatest loudness in the corresponding critical bands — this indicates the properties of the power-averaging employed, which gives more weight to louder transient energy than, for example, arithmetic or root-mean-square averaging. The verification comparisons demonstrate the close agreement with the reference results.

Specific roughness analyses for signal 2b are shown in Fig. 6. It can seen in Fig. 6a that the roughest parts of the signal are the bird calls, which feature rapid modulations of amplitude and frequency that both contribute to the roughness perception. These characteristics yield a time-dependent overall roughness $R(t)$ (not shown) that momentarily exceeds 0.2 asper$_{HMS}$, which is the threshold for time-aggregated roughness R deemed to indicate 'prominent' roughness advised by ECMA-418-2 [5]. This transient roughness behaviour is not discernible from the time-averaged results in Fig. 6b however, yet the spectral peak in the corresponding critical bands is observed. The specific roughness results are once again in very close

agreement with the reference.

Representative overall single-values calculated for each SQM are easy to interpret, as they reduce the analysis to a single dimension representing the overall magnitude of the SQ, facilitating quantitative comparisons between different sounds. Fig. 7 shows that the values provided by the implemented algorithms are in agreement with the reference results for all metrics.

4. CONCLUSIONS AND OUTLOOK

An implementation of the ECMA-418-2 SQMs for loudness, tonality and roughness has been developed and published as open-source software, licensed for open-access use. In the pursuit of ensuring robust verification testing, and to encourage wider application of the implemented algorithms, the code has also been incorporated into the SQAT toolbox [7, 8].

The results of rigorous verification testing consistently show that deviations between results calculated using the presented implementation compared with the reference are negligible for all audio signals examined, including both simple and complex sounds. The implemented algorithms are therefore considered to be fully verified. The verification exercise also highlighted the importance of verifying complex algorithms using a wide range of signals, as well as the value to developers of a database of reference results made available for verification purposes alongside a corresponding standard.

5. ACKNOWLEDGMENTS

MJBL acknowledges research funding by Horizon Europe and UK Research and Innovation (grant number 10061935). MJBL is also grateful to the UK Acoustics Network Plus for travel grant funding.

6. REFERENCES

[1] R. Sottek, *Modelle zur Signalverarbeitung im menschlichen Gehör (Models for signal processing in human hearing)*. Doctoral thesis, Rheinisch-Westfälische Technischen Hochschule Aachen (Rhenish-Westphalian Technical University of Aachen), 1993.

[2] R. Sottek, "A hearing model approach to time-varying loudness," *Acta Acustica united with Acustica*, vol. 102, no. 4, pp. 725–744, 2016.

[3] R. Sottek, "Calculating tonality of IT product sounds using a psychoacoustically-based model," in *Inter-noise Conference Proceedings*, vol. 250, pp. 5692–5703, Institute of Noise Control Engineering, 2015.

[4] R. Sottek, J. Becker, and T. Lobato, "Progress in roughness calculation," in *Inter-noise Conference Proceedings*, vol. 261, pp. 2835–2846, Institute of Noise Control Engineering, 2020.

[5] Ecma International, "ECMA-418-2:2024 Psychoacoustic metrics for ITT equipment — Part 2 (methods for describing human perception based on the Sottek Hearing Model)," International standard 3rd ed., 2024.

[6] M. J. B. Lotinga, "refmap-psychoacoustics." Computer software version 2025, 2025. https://github.com/acoustics-code-salford/refmap-psychoacoustics.

[7] G. Felix Greco, R. Merino-Martínez, A. Osses, and M. J. B. Lotinga, "SQAT: A sound quality analysis toolbox for MATLAB." Computer software version 1.3, 2025. https://doi.org/10.5281/zenodo.15148618.

[8] G. Felix Greco, R. Merino-Martínez, A. Osses, and S. C. Langer, "SQAT: A MATLAB-based toolbox for quantitative sound quality analysis," in *Inter-noise Conference Proceedings*, vol. 268, pp. 7172–7183, Institute of Noise Control Engineering.

[9] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and models*. Springer Series in Information Sciences, Berlin, Germany: Springer-Verlag, 3rd ed., 2007.

[10] R. Sottek, T. Lobato, M. Bender, and J. Becker, "Modeling the ISO 226:2023 equal-loudness-level contours by standardized loudness methods," European Acoustics Association, 2023.

[11] A. Fiebig and R. Sottek, "Contribution of peak events to overall loudness," *Acta Acustica united with Acustica*, vol. 101, no. 6, pp. 1116–1129, 2015.

[12] ISO, "ISO 532-1:2017 Acoustics — Methods for calculating loudness — Part 1: Zwicker method," international standard, 2017.

[13] M. C. Green and D. Murphy, "EigenScape: A database of spatial acoustic scene recordings," *Applied Sciences*, vol. 7, no. 11, p. 1204, 2017.

[14] M. J. B. Lotinga, M. C. Green, and A. J. Torija, "Human perception and response to sound from unmanned aircraft systems within ambient acoustic environments," *npj Acoustics*, vol. 1, no. 1, p. 2, 2025.