Variability of sound quality metrics for different aircraft types during landing and takeoff

Alves Vieira, Ana; Mehmood, Umair; Merino Martinez, Roberto; Snellen, Mirjam; Simons, Dick

DOI
10.2514/6.2019-2512

Publication date
2019

Document Version
Accepted author manuscript

Published in
25th AIAA/CEAS Aeroacoustics Conference

Variability of sound quality metrics for different aircraft types during landing and takeoff

Ana Vieira *, Umair Mehmood†, Roberto Merino-Martínez‡, Mirjam Snellen§, and Dick G. Simons¶
Section Aircraft Noise and Climate Effects, Faculty of Aerospace Engineering, Delft University of Technology, 2629 HS Delft, The Netherlands

The capacity of airports is limited due to the negative community response to noise. Traditional metrics, such as the A-weighted maximum sound pressure level ($L_{p,A,max}$), indicate the overall noise generated by an aircraft flyover but do not provide any information on tonal components or frequency variations in time that are known to affect annoyance. In this work 158 flyovers recorded at Amsterdam Schiphol Airport are analyzed in terms of sound quality metrics (SQM), loudness, roughness, tonality, sharpness and fluctuation strength. The recordings include landing and takeoff operations of 15 different aircraft types. The variability of the levels of the SQM are assessed per aircraft type. Possible correlations between the SQM and airframe and engine characteristics are investigated and empirical expressions for the loudness and roughness are formulated. The Effective Perceived Noise Level (EPNL) and the Psychoacoustic Annoyance Metric (PA_{mod}) are calculated for each flyover. The two metrics show a high correlation between them and this result is further investigated using listening tests. The listening tests show that tonality has a high importance in annoyance, however, its influence in PA_{mod} was found to be small.

I. Introduction

Communities close to airports experience high levels of noise for long periods of time, which has a negative impact on their health [1]. Considerable effort has been made by the aircraft industry towards more silent aircraft designs and over the last decades jet noise has been greatly reduced due to the increase of the bypass ratio of turbofan engines [2] and the installation of chevron mixers on the exhaust nozzles.

Traditional noise metrics, such as the A-weighted maximum sound pressure level ($L_{p,A,max}$) and the Effective Perceived Noise Level (EPNL), are indicators of the overall annoyance and allow to determine which aircraft or operating conditions produce more noise. However, such metrics do not provide information about what is behind such differences, and equal levels do not necessarily mean equal annoyance. The $L_{p,A,max}$ is a loudness based metric and EPNL is derived from noisiness curves and takes into account the presence of pure tones and the duration of the flyover [3]. These metrics were introduced in the 60s when aircraft noise characteristics were very different compared to modern aircraft.

Sound quality metrics (SQM) can be associated to different characteristics of noise, such as low or high frequency contents, tonal noise and fast or low frequency oscillations in time (for example buzzsaw noise during takeoff, which is generated when the velocity of the blade tips is supersonic) [4]. The five SQM considered (loudness, sharpness, fluctuation strength, tonality and roughness) provide information that can be used in aircraft design. In addition the SQM can be combined in a psychoacoustic annoyance metric (PA_{mod}), which gives a value of overall annoyance. The PA_{mod} metric is considered to provide a more accurate description of annoyance perceived by the human ear than EPNL [4].

This work analyzes 158 flyovers recorded at Amsterdam Airport Schiphol. The set of measurements includes 15 different types of aircraft, and landing and takeoff flyovers. The mean value and the variability of each SQM are investigated for each aircraft type. The correlations between the values of the SQM and aircraft characteristics, such as the wing span or engine bypass ratio, are investigated. This analysis aims to investigate whether the SQM can be associated to the aircraft configuration and if it is possible to find empirical expressions that relate them. Such empirical expressions would allow the design of less annoying aircraft in the perception of the residents, which does not necessarily imply an overall reduction of many decibels.

*PhD student, A.E.AlvesVieira@tudelft.nl.
†Consultant Aviation Policy at Ecorys NL
‡Postdoctoral researcher, R.MerinoMartinez@tudelft.nl.
§Associate Professor Acoustics
¶Chair Acoustics, Head Section Aircraft Noise and Climate Effects

The overall annoyance for each aircraft is determined using both EPNL and PA_{mod}. Listening tests are used to determine which metric is more in accordance with the perception of annoyance by the subjects. The listening tests are also used to investigate the weights attributed to the different SQM in the PA_{mod} and to propose improvements to these.

II. Theory

The five sound quality metrics employed (loudness, sharpness, roughness, tonality and fluctuation strength) are briefly described in this section and the methods used to calculate them are presented.

A. Loudness

Loudness is the subjective perception of the magnitude of a sound and it is dependent on its frequency, intensity and duration. Loudness has been standardized in ISO 532-1 [5] and it is expressed in phon (when in logarithmic scale) or sone (in linear scale).

The specific loudness N_s, which is the loudness in each critical frequency band z, is calculated using,

$$N_s(z) = 0.0635[10^{0.025L_{TQ}(z)}]\left[\left(0.75 + 0.25\left\{10^{0.1[L_E(z)-L_{TQ}(z)]}\right\}\right)^{0.25} - 1\right] \tag{1}$$

where L_E and L_{TQ} are the excitation level and the threshold in quiet conditions, respectively. The critical bands concept was introduced by Harvey Fletcher [6] and it is related to the neural activity of the human ear.

The values of specific loudness are determined in each critical band and it is checked whether they are masked by a sound concentrated in an adjacent critical band. The results are values of unmasked specific loudness in each critical band, N', which are used to determine the total loudness N_{total},

$$N_{total} = \int_0^{24} N'(z)dz. \tag{2}$$

B. Sharpness

Sharpness quantifies the high frequency content of a sound: a sound perceived as sharper has more high frequency content. This work uses the method of von Bismarck [7] to determine sharpness, S, given by

$$S = 0.11\,\frac{\int_0^{24} N'(z)g(z)dz}{N}. \tag{3}$$

Here $g(z)$ is a weighting function, given by

$$g(z) = \begin{cases} 1 & z \leq 16 \\ 0.066e^{0.171z} & z > 16. \end{cases} \tag{4}$$

This weighting function causes the value of sharpness to be higher for higher critical bands.

C. Tonality

Tonality measures the tonal prominence of a sound and it was first developed by Terhardt [8]. This method is based on the virtual pitch theory, which states that the first six to eight harmonics of a complex tone can be perceived as separate spectral pitches.

The tonal components are extracted from the signal and the corresponding spectral samples are investigated, where it is tested if the sound pressure level (SPL or L_p) value of the i^{th} sample, L_i, is higher than the next lower (i-1) sample and the next higher (i+1) sample. Then it is verified if the candidate sample i meets the condition,

$$L_i - L_{i+m} \geq 7\,\text{dB} \quad \text{for} \quad m = -3, -2, .., 2, 3. \tag{5}$$

The next step of the method evaluates the masking effects. The SPL excess ΔL_i is calculated using

$$\Delta L_i = L_i - 10\log_{10}\left\{\left[\sum_{k\neq i}^{n} A_{Ek}(f_i)\right]^2 + E_{Gr}(f_i) + E_{HS}(f_i)\right\}. \tag{6}$$

Here n is the number of tonal components, $A_{\mathrm{E}k}$ is the amplitude of the secondary neural excitation at frequency f_i due to the k^{th} tonal component. E_{Gr} is the masking intensity of the broadband noise encompassing the tone and E_{HS} is the intensity at the threshold of hearing. Expressions for those terms can be found in [8].

Equation (6) is used to calculate ΔL_i for all tonal components from $i = 1$ to $i = n$. Positive values of ΔL_i indicate components aurally relevant. The interaction of simultaneous spectral components in the auditory system means that the spectral pitches are not exactly the same as the pitches of isolated tones with the same frequency. The individual spectral pitch of a tonal component can be determined using

$$H_i = f_i(1 + \nu_i), \tag{7}$$

in which ν_i is an expression dependent of variables calculated in previous steps of the procedure, taking into account the pitch shifts. These would be ignored if H_i was considered equal to f_i.

The last step of the method consists in weighting the different tonal components as,

$$\mathrm{WS}_i = \left[1 - e^{\frac{\Delta L_i}{15}}\right]\left[1 + 0.07\left(\frac{f_i}{0.7} - \frac{0.7}{f_i}\right)\right]^{-0.5}, \tag{8}$$

and find $H_{im}(f_i, \nu_i)$, which is the virtual pitch of the m^{th} sub-harmonics of the i^{th} relevant component, given by

$$H_{im} = \frac{f_i}{m}\left\{1 + \nu_i - 0.01\mathrm{sign}(m-1)\left[18 + 2.5m - (50 - 7m)\frac{f_i}{m} + 0.1(\frac{f_i}{m})^{-2}\right]\right\}. \tag{9}$$

This method, designated as Terhardt's method, can only be applied for pitch values H_{im} lower than 500 pitch units (pu) and does not result in a single value of tonality, which is useful when evaluating different sounds. In addition, it was found that the precision of tonality can increase when loudness is introduced in the calculation [9].

Those drawbacks are eliminated when using the Aures' tonality metric, which has been widely used in the automotive industry. Recent studies show that it is also adequate to aircraft noise [4]. In this Aures' tonality metric, the weighting of the tonal components includes a loudness term, and tonality is expressed as

$$K = C\, W_T^{0.29} W_{\mathrm{Gr}}^{0.79}. \tag{10}$$

Where W_{Gr} is the tonal loudness weighting, which is the relation between the total loudness of a sound and the loudness excluding the tones and W_T is tonal weighting factor, both introduced by Aures. c is a calibration term to give a 1 kHz tone of 60 dB SPL a tonality of one.

D. Roughness

Roughness assesses fast loudness fluctuations (between 50 to 90 Hz). This work uses the method of Zwicker and Fastl to estimate roughness [10]. It was found that two characteristics of the ear influence the roughness perception: the frequency selectivity of the hearing system at low frequencies, and the limited temporal resolution at high frequencies. The model of roughness proposed by Zwicker and Fastl uses the temporal masking pattern. Fig. 1 shows the relation between a masker and its temporal masking pattern.

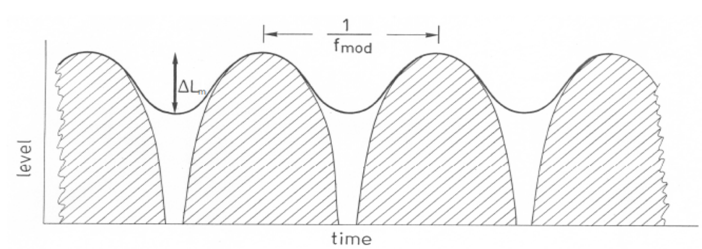


Fig. 1 Scheme of temporal masking used by Zwicker and Fastl [10].

In Fig. 1 f_{mod} is the modulation frequency, which is the interval between two consecutive peaks of the masker envelope.

According to Zwicker and Fastl model roughness is given by,

$$R = 0.3 f_{\text{mod}} \int_0^{24} \Delta L_m(z) dz. \tag{11}$$

Here z is the critical band rate, f_{mod} is the modulation frequency and $\Delta L_m(z)$ is the modulation depth of the specific loudness at the critical band z.

The calculation of f_{mod} can be challenging for complex signals and its value can change for different critical bands. The best approch to combine different modulations is still object of research. In this work the model of Daniel & Weber [11], based on the Aures' model, is used.

E. Fluctuation strength

Fluctuation strength assesses slow fluctuations in loudness, and its value is maximum for loudness fluctuations around 4 Hz. Zwicker and Fastl propose two distinct methods for calculating fluctuation strength, one for tonal noise and the other for broadband noise [10].

The fluctuation strength, FS, of a sinusoidally amplitude-modulated broadband noise is given by

$$FS = \frac{5.8(1.25 m_d - 0.25)(0.05 L_{\text{BBN}} - 1)}{(f_{\text{mod}}/5)^2 + (4/f_{\text{mod}}) + 1.5}. \tag{12}$$

Here m_d is the modulation depth and L_{BBN} is the level of broadband noise.

The modulated tones depend on the frequency, so instead of one masking depth as in Eq. 12, all the masking depths are integrated along the critical band, and the FS of tonal components can be approximated as,

$$FS = \frac{0.008 \int_0^{24} \Delta L dz}{(f_{\text{mod}}/4) + (4/f_{\text{mod}})}. \tag{13}$$

F. Psychoacoustic annoyance

Aircraft noise is not steady in time, and the SQM change during the flyover. For that reason, in this work the value of the SQM that is exceeded 5% of the time is considered. Therefore, when loudness (N_5), tonality (K_5), sharpness (S_5), roughness (R_5) and fluctuation strengh (FS_5) are quantified in this work, they refer to the value exceeded 5% of the time.

The five SQM can be combined to obtain a single value to quantify annoyance. Zwicker and Fastl introduced a psychoacoustic annoyance model (PA) [10], expressed by

$$PA = N_5 \left(1 + \sqrt{\omega_S^2 + \omega_{FS}^2}\right), \tag{14}$$

in which the terms ω_S and ω_{FS} are determined by Eq. 15 and Eq. 16, respectively,

$$\omega_S = \begin{cases} 0.25(S - 1.75)\log_{10}(N_5 + 10), & S > 1.75 \\ 0 & , S < 1.75 \end{cases} \tag{15}$$

$$\omega_{FS} = \frac{2.18}{N_5^{0.4}}(0.4 FS + 0.6 R). \tag{16}$$

This model was modified by More [4] to include tonality, since it is important in aircraft noise. The modified psychoacoustic annoyance model (PA_{mod}) is expressed as

$$PA_{\text{mod}} = N_5 \left(1 + \sqrt{\gamma_0 + \gamma_1 \omega_S^2 + \gamma_2 \omega_{FS}^2 + \gamma_3 \omega_T^2}\right). \tag{17}$$

The variables γ are constants and ω_T^2 is the new term that includes the tonality, and it is given by

$$\omega_T^2 = \left[\left(1 - e^{-\gamma_4 N_5}\right)^2 \left(1 - e^{-\gamma_5 K_5}\right)^2\right]. \tag{18}$$

III. Experimental Setup

Flyover measurements of different aircraft types were recorded at Amsterdam Airport Schiphol during three days with similar weather conditions. The measurement system, see Fig. 2, consists of an acoustic array of 64 microphones distributed in an Underbrink spiral configuration [12], and has the dimensions of 4 m × 4 m × 0.12 m. The microphones of the array can be used collectively to identify different noise sources using beamforming and apply the SQM to the different elements [13].

Fig. 2 Acoustic camera used to record the flyovers at Schiphol Airport.

The wooden plates that compose the array are covered with acoustic absorbing foam to avoid reflections. The foam selected was Flamex GU of 15 mm due to its high absorption coefficient. The acoustic array has adjustable height legs in order to compensate for irregularities of the floor.

The microphone (PUI AUDIO 665-POM-2735P-R [14]) signals were sampled at 50 kHz. Also an optical camera was used (Datavision UI-1220LE with a lens Kowa LM4NCL), which had a frame rate of 30 Hz.

The type of aircraft and its height and velocity were determined using an ADS-B (Automatic Dependent Surveillance-Broadcast) receiver. Since not all aircraft have an ADS-B transponder, the aircraft type was also verified with online flight trackers and consecutive frames of the optical camera were used to check the height and velocity.

The microphone array was positioned at an extension of runway 18C of Schiphol airport. This choice was due to the existence of a public open space close to that runway where it was possible to place the array. In this extension of the runway the flyover altitudes overhead during landing are around 60 m. This track is mainly used for landing aircraft, but not exclusively.

Table 1 shows the number of landing flyovers recorded per aircraft type and Table 2 displays the number of takeoff flyovers, which are considerably fewer.

Table 1 Landing flyovers recorded in Schiphol Airport.

Aircraft	Nº of flyovers	Aircraft	Nº of flyovers
A319	4	B737	63
A320	13	B777	9
A321	3	B787	8
A380	2	CRJ-900	2
Avro RJ85	3	Fokker 70	7
ERJ-175	15	ERJ-190	22

Table 2 Takeoff flyovers recorded in Schiphol Airport.

Aircraft	Nº of flyovers	Aircraft	Nº of flyovers
A320	3	B737	2
A321	1	ERJ-190	1

The same duration of the flyovers (10 s) was used for the calculation of the SQM, EPNL and psychoacoustic annoyance metrics in order to obtain comparable results.

IV. Results

A. Assessment of sound quality metrics for flyover measurements

The flyovers are analyzed in terms of SQM and their variability within the same aircraft type. The analysis is presented only for landing flyovers due to the limited number of available takeoff flyovers. The latter are only used to verify the expected differences in terms of SQM between takeoff and landing flyovers.

The flight trajectories during final approach are more regular than those at takeoff [15], so no great variations are expected for the SQM. The flight trajectory and aircraft settings affect the values of the SQM, and therefore the variability of such parameters should be accounted for. The variability of absolute altitude, total ground speed and the rotational speed of the low-pressure shaft on which the fan is mounted (N_1) [16, 17] of the landing aircraft are displayed in Figs. 3 to 5, respectively. All the values were calculated for the overhead time and the aircraft types are sorted in ascending order of their maximum take-off weight.

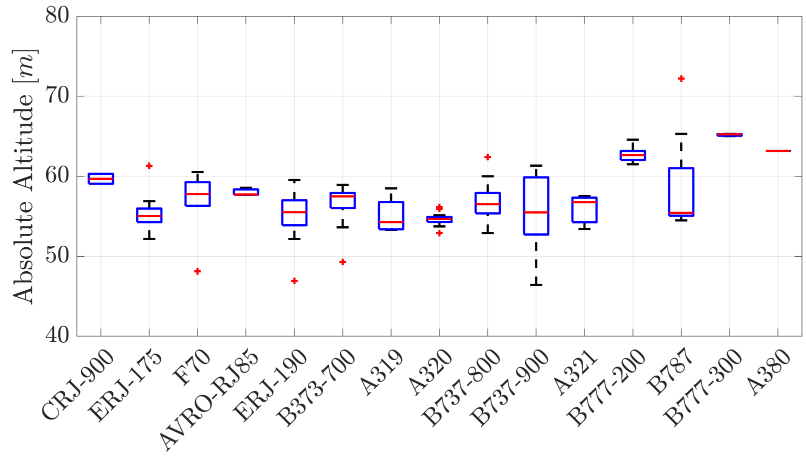


Fig. 3 Absolute altitude variability for landing flyovers.

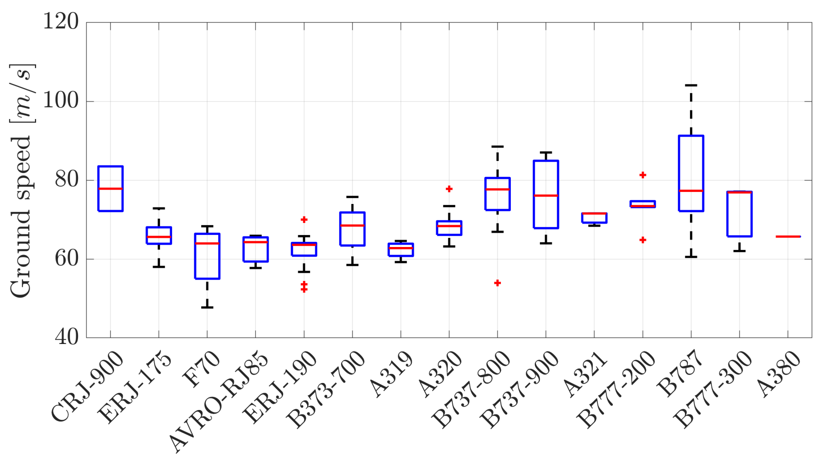


Fig. 4 Ground velocity variability for landing flyovers.

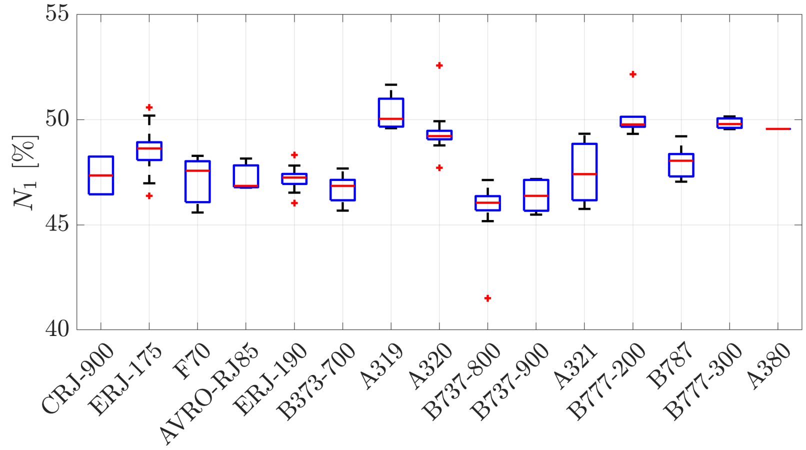


Fig. 5 N_1 variability for landing flyovers.

The altitude of the aircraft relative to the ground during overhead is between 50 m to 60 m for all aircraft except the B777 and the A380, which are at a higher altitude. The altitude does not present much variability within the same aircraft type. Only the B737-900 and the B787 show significant variability. The same tendency is observed for the ground velocity, with the higher variability verified for the same aircraft types. The average values of ground speed values vary between 60 m/s and 80 m/s. The values of N_1, obtained following the method of Schluter and Becker [18], vary between 45 % and 50 % for most aircraft, with the A321 and F70 presenting the higher variations.

The variability of the flight trajectory and fan settings observed in Figs. 3 to Fig. 5 between aircraft types is relatively low, meaning that it can be assumed that significant differences observed in the SQM are not associated with a drastic difference in the flight trajectory and operating conditions, but to differences in aircraft design.

Figure 6 to Figure 10 show the SQM of each aircraft type. All the SQM are calculated at the observer position (microphone array). Figure 6 shows the results of loudness determined for the different aircraft recorded during landing. The aircraft with higher loudness are the B777 series and the A380, as expected because these are the two aircraft with larger dimensions and thrust.

Two other aircraft stand out but due to their low value of loudness, the F70 and the CRJ900. These two models have aft-mounted engines and therefore engine noise is partially shielded by the wings and the fuselage of the aircraft [19].

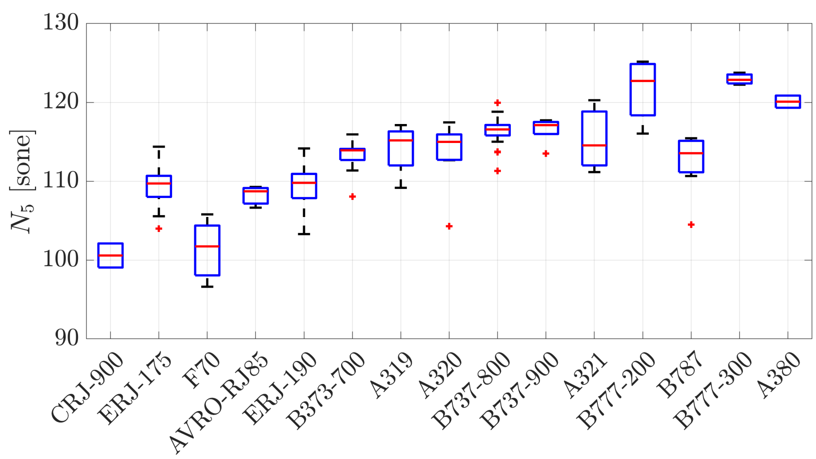


Fig. 6 Loudness levels and variability for landing flyovers.

The values of sharpness for each aircraft are shown in Fig. 7 and they show more variation (relative to the maximum value) within the same aircraft type than loudness. The same approximate value of sharpness was found for all the aircraft (≈ 2.2 acum). The exceptions are the B787 and the Avro-RJ85, which present the lowest and highest values of sharpness, respectively. High frequency content can be associated to the fan, compressor and turbine due to the tones. In the Avro-RJ85 the four engines result in a higher value of sharpness compared with all the other aircraft types, which are twin-engined, except the A380, but the engines are widely separated along the wing span on the contrary of the Avro-RJ85.

Figure 8 displays the values of tonality for all the aircraft types. The tonal content is often related with the fan rotational speed and tones in the airframe (nose landing gear). The aircraft B777-200, A319, A320 and ERJ-175 present the higher values of tonality. However, the comparison of Fig. 8 and Fig. 5 shows that there is not a correlation between N_1 and K_5, i.e., higher values of N_1 do not correspond to higher values of tonality.

The variability of roughness is shown in Fig. 9. Roughness is expected to be more prominent during takeoff than landing, due to buzzsaw noise. In this set of landing aircraft the value of roughness remains approximately the same (≈ 0.1 asper) and it is the SQM with the smallest variations within the same aircraft model. The larger aircraft as the B777 and the A380 stand out for their high values of roughness compared with other aircraft types.

The fluctuation strength is represented in Fig. 10 for completeness. However, it is not an important metric in aircraft noise. The perception of low frequency oscillations is associated with the wind and background noise. Fig. 10 indicates a large variability of the fluctuation strength and approximately the same average value for all aircraft, indicating that indeed it is not related with the aircraft type, but to external conditions.

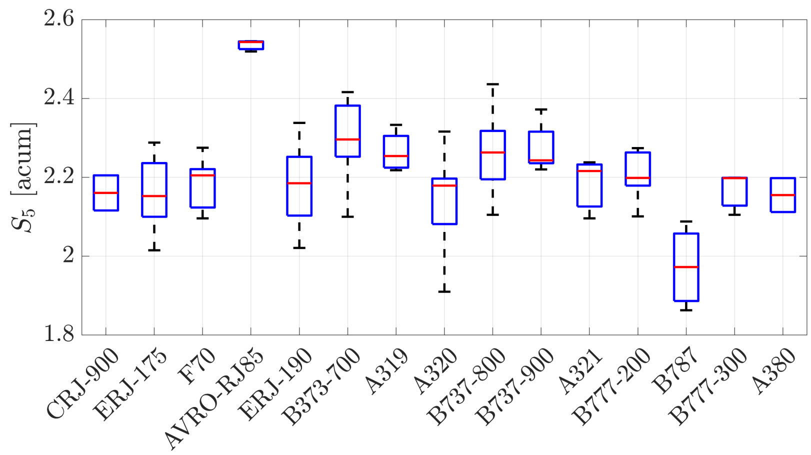


Fig. 7 Sharpness variability for landing flyovers.

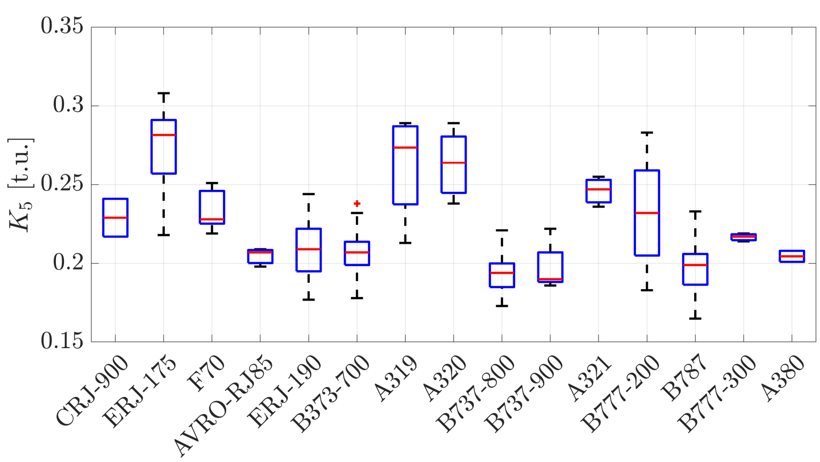


Fig. 8 Tonality variability for landing flyovers.

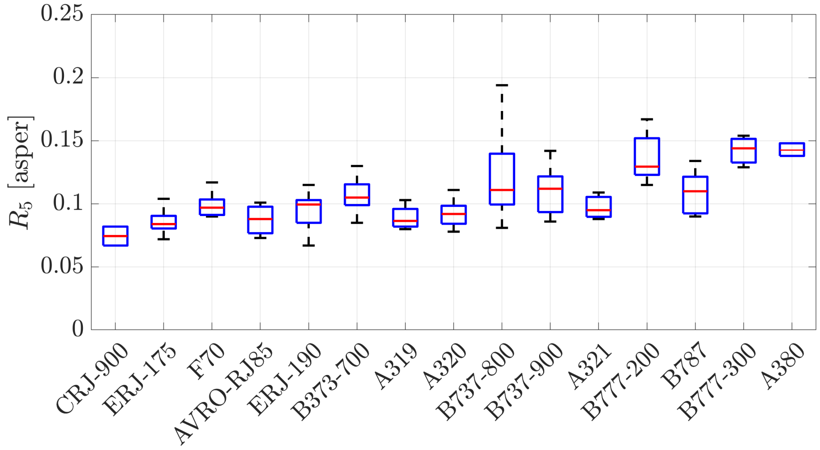


Fig. 9 Roughness variability for landing flyovers.

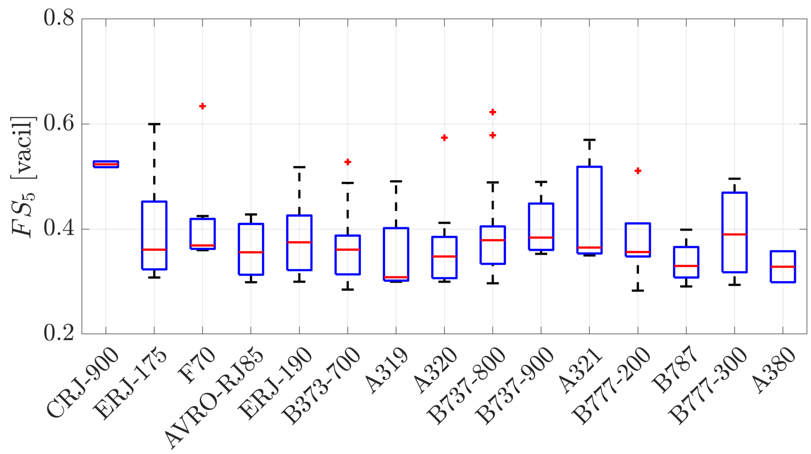


Fig. 10 Fluctuation strength variability for landing flyovers.

The plots of Figure 11 compares the average values of SQM for 5 different aircraft during landing and take-off. The fluctuation strength was excluded as it is not of importance for aircraft noise. Loudness is considerably lower for all aircraft during take-off. As the engines are at full power during takeoff, one would expected higher values of loudness, however, the height of landing and takeoff differ in the same measuring position. The estimated height for all landing aircraft is around 60 m but for takeoff it is more variable and exhibits values around 300 m.

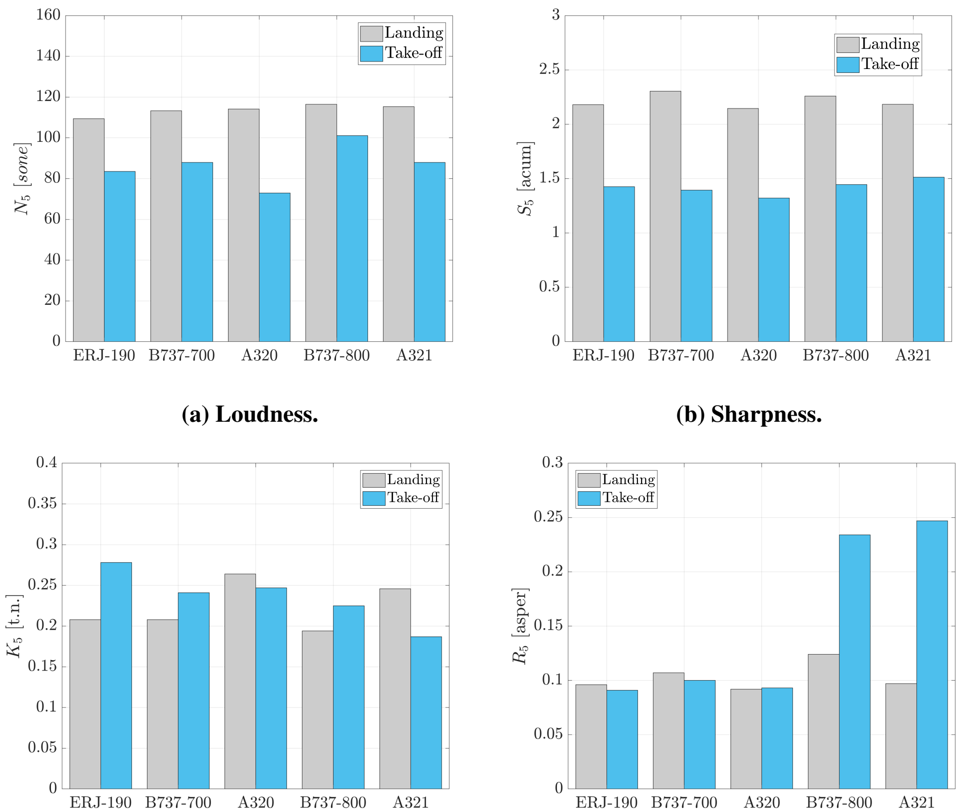
The value of sharpness, S_5, is higher during landing for the five aircraft. The values of S_5 during take-off are so low that they are neglected in the calculation of psychoacoustic metrics (see Eq. (15)). This result was expected because sharpness is a measure of the high frequency content of aircraft noise, which is lower during take-off due to a strong presence of jet noise (low frequency) that masks the high frequency content.

Tonality is expected to be higher during takeoff than during landing due to buzzsaw tones. However, that is only verified for three of the five aircraft analyzed in Figure 11c, the B737-700, B737-800 and ERJ-190.

The roughness is also associated with buzzsaw noise, due to high speed at the blade tip of the fan. For the aircraft analyzed, roughness is higher during takeoff or approximately the same as during landing, except for the A321 and the B737-800.

The available data indicates that the SQM differ according to the flight phase, despite the limited data available of take-off flyovers:

- Sharpness during take-off is very low compared with landing because the high frequency content is masked by jet noise,



(a) Loudness. **(b) Sharpness.**

(c) Tonality. **(d) Roughness.**

Fig. 11 Difference sound quality metrics per aircraft type during landing and take-off.

- Roughness tends to be higher during take-off due to fast frequency oscillations generated by buzzsaw noise,
- Tonality is not clearly higher during take-off, although it is expected to be higher, due to the higher value of N_1, the tones are masked by jet noise,
- Loudness and roughness are seen to be directly correlated with the dimension of the aircraft.

B. Correlation of the Sound Quality Metrics with aircraft characteristics

Some SQM can be directly associated to the aircraft characteristics, e.g. the very low value of loudness of the F70 and CRJ-900 compared with other aircraft types is hypothesized to be due to shielding of engine noise. This subsection explores possible correlations between the SQM and aircraft characteristics. For that, the average value of the SQM for each aircraft type is used.

Different characteristics of the aircraft fuselage and engines were selected to find possible correlations with the SQM. The airframe characteristics selected were the wing span, length of aircraft, maximum take-off weight (MTOW), cabin diameter, number of tires of the nose and main landing gear and respective diameter. The parameters selected for the engine were the fan diameter, the bypass ratio (BPR) and the number of fan blades. These characteristics were selected based on their availability in literature.

No significant correlation was found for the tonality, K_5 with any of the airframe or engine parameters. The correlation with FS_5 was not investigated as this metric is not related with characteristics of the aircraft.

The values of the coefficient of determination (R^2) and the p-value for each correlation tested are displayed in Table 3. The correlations with no significance are represented by a dash symbol in the table and parameters referred before but that showed no correlations with the metrics were not included in the table. Despite the reduced number of takeoff flyovers available, they are also included in the analysis. The results obtained for the takeoff flyovers should be interpreted with care and cannot be considered as conclusive.

Table 3 Values of R^2 and p-value (between parenthesis) for different correlations between the SQM and the aircraft and engine characteristics.

	N_5		R_5		S_5	
	Landing	TO	Landing	TO	Landing	TO
Wing span	0.48	-	0.72	-	0.44	-
	(4.45e-3)	-	(6.73e-5)	-	(2.23e-1)	-
Fuselage length	0.44	-	0.69	-	-	-
	(7.20e-3)	-	(1.20e-4)	-	-	-
Cabin diameter	0.60	-	0.73	-	-	-
	(3.0e-3)	-	(4.72e-5)	-	-	-
MTOW	0.39	-	0.64	-	-	-
	(1.30e-2)	-	(3.70e-4)	-	-	-
MLG tire diameter	0.63	-	0.43	-	-	-
	(3.80e-4)	-	(8.30e-3)	-	-	-
Fan diameter	0.48	-	0.68	-	-	-
	(4.1e-3)	-	(1.4e-4)	-	-	-
BPR	0.43	-	0.47	-	-	0.90
	(9.8e-2)	-	(4.75e-3)	-	-	(1.30e-2)
Number of fan blades	-	0.47	0.27	-	0.33	0.73
	-	(2.02e-1)	(4.66e-2)	-	(2.32e-2)	(6.39e-2)

The conclusion to be drawn is that there are no correlations between the SQM and the fuselage characteristics during this flight phase. This result was expected as engine noise is more significant than airframe noise during takeoff. On the other hand, strong correlations were found between sharpness and the engine parameters during take-off, even though sharpness has low values during take-off, as seen in the previous subsection.

As expected loudness presents a correlation with all the fuselage parameters during landing, as airframe noise plays a significant role and larger surfaces result in higher levels of noise.

Roughness also shows a correlation with most of the selected parameters, during landing. Roughness is associated with fast amplitude modulations of a sound, and therefore it is expected to be more significant during takeoff due to buzz-saw noise. During landing, buzz-saw noise does not play a role, so other sources of noise, as the fairings, wheels and hoses, become important [20, 21]. Some of these details of the aircraft structure are responsible for low frequency noise and, therefore, an aircraft of larger dimensions results in higher values of roughness (see Fig. 9, where the B777 and the A380 present the higher values of R_5).

Several significant correlations were found both for landing and takeoff, indicating that the SQM can be used during the design phase of an aircraft to reduce the annoyance on ground. However, not all results were as expected, such as the non-existence of significant correlations between K_5 and any fan characteristics.

This works makes a first attempt to find empirical expressions that relate the SQM with the aircraft characteristics. Such empirical expressions would be of use in the preliminary design phase of conventional aircraft. Due to the results of Table 3, only expressions for N_5 and R_5 during landing are investigated.

A linear regression was used to find the coefficients matrix γ, using the 158 flyovers. The fuselage and engine parameters showing higher correlation with the SQM (i.e., higher value of R^2 in Table 3) were used to find the coefficients γ and then discarded when very small compared with the weight that other aircraft parameters played in the expression. This process resulted in the empirical expressions of Eq. 19 and Eq. 20, for the loudness and roughness during landing, respectively,

$$N_5 = \gamma_{0,N_5} + \gamma_{1,N_5} D_{\text{MLG}} + \gamma_{2,N_5} D_{\text{cabin}} + \gamma_{3,N_5} \text{BPR}, \tag{19}$$

$$R_5 = \gamma_{0,R_5} + \gamma_{1,R_5} D_{\text{cabin}} + \gamma_{2,R_5} N_{\text{blades}} + \gamma_{3,R_5} \text{BPR}. \tag{20}$$

Where D_{cabin} is the diameter of the cabin, D_{MLG} is the diameter of the tires of the main landing gear, N_{blades} is the number of the blades of the fan and BPR is the bypass ratio.

The values of the coefficients γ obtained for the empirical expressions are displayed in Table 4. Also the correlation between the results obtained by the empirical expressions and the experimental data is presented.

Table 4 Coefficients of the empirical expression for loudness and roughness.

	γ_0	γ_1	γ_2	γ_3	R^2	p-value
N_5	81.189	0.730	25.804	0.167	0.668	2e-4
R_5	0.125	0.007	-0.002	0.002	0.791	9e-6

The correlation between experimental and predicted values of loudness and roughness is presented in Fig. 12 and Fig. 13, respectively. For the loudness correlation, only two aircraft show a large discrepancy with the prediction, the F70 and the B787. The F70 presents engine noise shielding, which is not quantified in the empirical expression, therefore the value of loudness is lower than expected. The B787 does not present a behavior similar to aircraft of equivalent size, as seen in Section V.A, showing lower values of loudness, sharpness and tonality, so it does not fit the curve. For the roughness curve, all aircraft show a strong correlation with the predicted value except the CRJ-900.

The investigation of empirical expressions for the SQM require more experimental data and aircraft types with more distinct characteristics, however, this first study shows encouraging results. More complex expressions employing exponentials and polynomials should be investigated in future work. More data is required to draw any conclusions about the takeoff correlations.

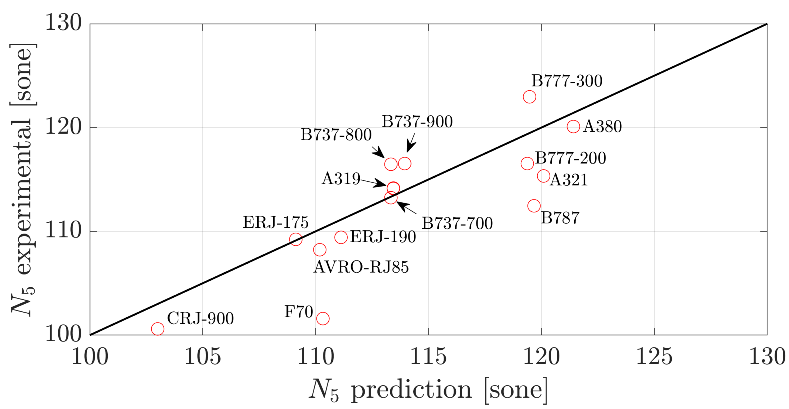


Fig. 12 Experimental and predicted (using Eq. 19) values of roughness. The black line indicates a correlation of $R^2 = 1$ and the red dots are the experimental values. The experimental and predicted value is more in agreement as closer to the black line.

C. Comparison of EPNL with PA_{mod} and modification of the metric

The SQM are combined to calculate average values of PA_{mod} for each aircraft type. The average value of EPNL for each aircraft type is also calculated and compared with the corresponding PA_{mod} value. Figure 14 displays the average value of EPNL and PA_{mod} for each aircraft type normalized by the maximum value, i.e. the maximum value found for all flyovers. Only landing flyovers are considered in the plot. The B777-200 and B777-300 were found to be the most annoying aircraft both by the EPNL and the PA_{mod}. The EPNL clearly considers the A380 as approximately annoying as the B777, but PA_{mod} considers that the B737 series and the A319 are more annoying than the A380. The two metrics agree in which aircraft are the least annoying, the F70 and the CRJ-900.

The B787 appears as one of the most silent aircraft both using EPNL and PA_{mod}, which was not expected due to its dimension. However, this aircraft type is relatively new so it is equipped with state-of-the-art noise reduction devices, such as chevrons. This aircraft shows a low value of sharpness and tonality (Fig. 7 and Fig. 8) when compared with other aircraft, which can explain the low values of EPNL and PA_{mod}.

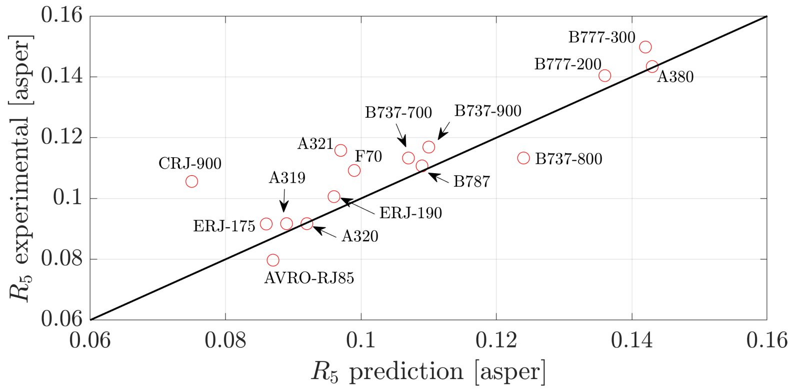


Fig. 13 Experimental and predicted (using Eq. 20) values of roughness. The black line indicates a correlation of $R^2 = 1$ and the red dots are the experimental values. The experimental and predicted value is more in agreement as closer to the black line.

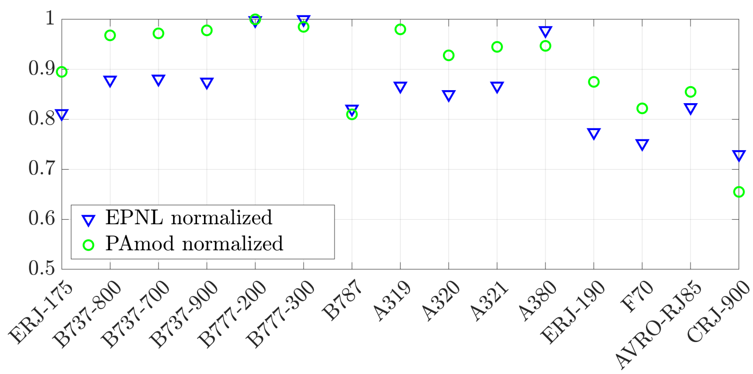


Fig. 14 Average values of EPNL and P_{mod} for each aircraft. The values are normalized by the maximum value of EPNL and P_{mod} of all aircraft analysed.

The normalized values of EPNL and PA_{mod} show similar trends for most aircraft, despite the differences highlighted before, with a coefficient of determination R^2=0.53 between them.

Listening tests were performed to evaluate which metric was more in agreement with perceived annoyance. These tests were conducted in the anechoic room of the Faculty of Applied Sciences of Delft University of Technology. A group of twenty subjects (male and female, aged from 21 to 61 years old) participated in the experiment. These subjects had to evaluate ten different sounds of the data set used in this work, using a scale of 0 to 100, which was presented to them as in Fig. 15.

Fig. 15 Scale used in the subjective tests.

The sounds were played by two loudspeakers positioned relative to the subjects as represented in Fig. 16. The distances between the walls, observers and loudspeakers were chosen following the work of Hoeg et al [22].

The correlation between the direct rating of the subjects and the values of EPNL and PA_{mod} calculated for the 10 flyovers is displayed in Fig. 17. The subjective rating of each flyover is an average of the values attributed by the 20 subjects. Both EPNL and PA_{mod} are in agreement with the subjective ratings, with PA_{mod} presenting a slightly better

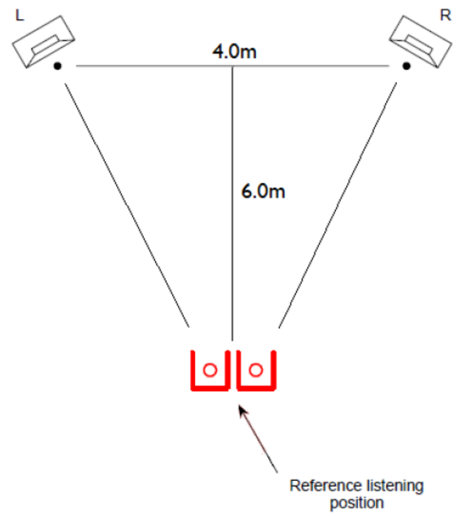


Fig. 16 Setup used in the subjective tests.

correlation. To further investigate this, the SQM are also presented for all audio files in Table 5.

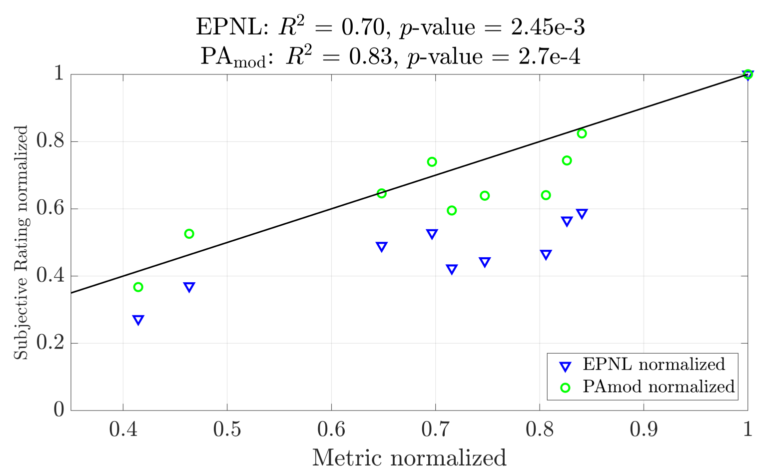


Fig. 17 Correlation between the direct rating of the subjective tests and the values of EPNL and P$_{mod}$. The black line indicates a correlation of $R^2 = 1$.

In Table 6 the sound files of Table 5 are ordered from the most to the least annoying in the first column, according to the subjective ratings attributed by the subjects. In the second column the sound files are ordered from the most to the least annoying according to the EPNL and in the third column according to the PA_{mod} results.

The sound files #8 and #10 are considered the most annoying, which is predicted correctly both by the EPNL and PA_{mod}. The two least annoying sound files, #1 and #4 are also correctly predicted by the two metrics. The EPNL is also capable of predicting correctly the third most annoying sound file, #7, and PA_{mod} the third least annoying sound, #6. However, the sound files in the middle of the table are not in the order attributed by the subjects, using EPNL or PA_{mod}.

The three most annoying sounds according to the subjective ratings are those with higher value of loudness, N_5, according to Table 5. The two least annoying sounds, #4 and #1 correspond to the sound files with lower value of N_5. When differences in the value of N_5 are not so evident, the metrics show more difficulties in predicting the order attributed by the subjects. The sound files #9, #5 and #6 are considered more annoying than #3, even though #3 has a higher value of N_5. However, #9, #5 and #6 present a high values of tonality, K_5, when compared with #3.

The importance of tonality has been emphasized in previous works and More [4] introduced a tonal term in the

Table 5 Sound quality metrics calculated for the sound files used in the subjective tests.

Sound file	N_5	K_5	S_5	R_5	FS_5	EPNL	PA_{mod}
#1	112.109	0.255	1.928	0.075	0.372	101.310	167.220
#2	128.880	0.232	2.051	0.088	0.473	105.360	205.380
#3	140.583	0.225	2.186	0.085	0.325	106.430	235.250
#4	80.781	0.219	1.567	0.197	0.592	96.900	116.840
#5	132.143	0.240	2.057	0.075	0.390	103.950	203.230
#6	123.320	0.262	2.063	0.086	0.372	103.240	189.240
#7	147.254	0.224	2.090	0.107	0.304	107.420	236.510
#8	215.345	0.310	1.722	0.173	0.542	115.630	318.020
#9	124.786	0.247	2.162	0.080	0.468	104.650	203.740
#10	158.831	0.228	2.219	0.103	0.364	107.990	262.060

Table 6 Sound files used in the experiment order from the most to least annoying according to subjective ratings, EPNL and PA_{mod}.

Subj. ratings	EPNL	PA_{mod}
#8	#8	#8
#10	#10	#10
#7	#7	#3
#9	#3	#7
#5	#2	#9
#6	#9	#5
#3	#5	#2
#2	#6	#6
#1	#1	#1
#4	#4	#4

original Psychoacoustic Annoyance model (see Eq. 17). However, this new term does not contribute significantly to the final value of PA_{mod} in the set of sound files analyzed in this work. In Fig. 18 the values of the SQM used are between the minimum and maximum values estimated for the sound files. Only one SQM is varied for each curve, and all the others remain fixed. The fixed SQM are the ones corresponding to sound file #5, because it is located at the centre of Table 6 and, therefore, it is a sound of average annoyance. Figure 18 shows that PA_{mod} mostly depends on loudness and sharpness. The effect of roughness and fluctuation strength of this set of sounds in PA_{mod} are negligible and the influence of tonality is small, which explains why the sounds with average annoyance are not in agreement with the subjective ratings.

V. Conclusions and Future Work

This work analyzed the sound quality metrics of aircraft under operational conditions. The 158 flyover measurements of 15 aircraft types were used to calculate the loudness, tonality, roughness and sharpness associated with each aircraft model.

The sound quality metrics were found to differ for landing and take-off aircraft, which was expected. Correlations of the metrics with characteristics of the airframe and engines were investigated. Due to the limited number of take-off

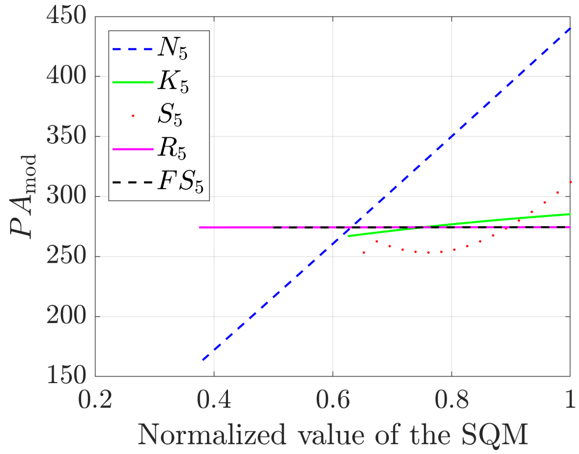


Fig. 18 Variation of PA_{mod} with the SQM in the range of the sound files used in the subjective tests.

flyover measurements the work focused mainly on landing flyovers, which showed a strong correlation of loudness and roughness with characteristics of the fuselage such as the cabin diameter and the diameter of the tires of the main landing gear. Empirical expressions relating the roughness and loudness with the characteristics of the aircraft were obtained using the experimental data. A good agreement was verified between the results of loudness and roughness from the empirical expression exists for each aircraft type with the experimental values.

The sound quality metrics of each aircraft type were combined in the psychoacoustics annoyance model, which assesses the total annoyance of a signal. Those values were compared with the EPNL, and it was found that the two metrics were not in agreement for all aircraft types analyzed.

Listening tests were performed to investigate differences between the EPNL and the PA_{mod} and it was verified that both metrics did not have difficulties in identifying the most and least annoying sounds according to the listenning tests, because of the distinctive values of loudness. It was verified that a good agreement between the listening tests and both the EPNL and the PA_{mod} exists, but slightly better for the PA_{mod}. For sounds with similar loudness, both metrics do not show a good agreement with the listening tests. For the case of the PA_{mod} it was found that the reason of such discrepancy is the low influence of the tonality in the final value of annoyance.

For future work, more experimental data of flyovers and listening tests is required. Additional flyover data of landing and take-off aircraft and of more aircraft models can be used to improve the preliminary empirical expressions for the sound quality metrics found in this work. In addition, a larger dataset of subjective ratings of flyovers can be used to investigate an alternative expression for PA_{mod} in which the tonality has a higher influence in the final value of annoyance.

References

[1] Basner, M., and McGuire, S., “WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Effects on Sleep,” *International Journal of Environmental Research and Public Health*, Vol. 15, No. 3, 2018. doi:10.3390/ijerph15030519.

[2] Papamoschou, D., “New method for jet noise reduction in turbofan engines,” *AIAA journal*, Vol. 42, No. 11, 2004, pp. 2245–2253.

[3] Peppin, R. J., *Community Noise*, ASTM International, 1979.

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

[5] ISO532-1, *ISO 532-1 (2017). Acoustics – Method for calculating loudness – Zwicker method*, Standard (International Organization for Standardization, 2017.

[6] Allen, J. B., “Harvey Fletcher’s role in the creation of communication acoustics,” *The Journal of the Acoustical Society of America*, Vol. 99, No. 4, 1996, pp. 1825–1839.

[7] von Bismarck, G., "Sharpness as an attribute of the timbre of steady sounds," *Acta Acustica united with Acustica*, Vol. 30, No. 3, 1974, pp. 159–172.

[8] Terhardt, E., Stoll, G., and Seewann, M., "Algorithm for extraction of pitch and pitch salience from complex tonal signals," *The Journal of the Acoustical Society of America*, Vol. 71, No. 3, 1982, pp. 679–688. doi:10.1121/1.387544.

[9] Shrestha, M., and Zhong, Z., "Sound quality user-defined cursor reading control-tonality metric," Master's thesis, Technical University of Denmark, DTU, DK-2800 Kgs. Lyngby, Denmark, 2003.

[10] Fastl, H., and Zwicker, E., *Psychoacoustics: facts and models*, 3rd ed., No. 22 in Springer series in information sciences, Springer, Berlin ; New York, 2007.

[11] Daniel, P., and Weber, R., "Psychoacoustic roughness: Implementation of an optimized model," *Acustica*, Vol. 8, No. 113, 1997.

[12] Underbrink, J. R., "Circularly symmetric, zero redundancy, planar array having broad frequency range applications," , U.S. Patent number 6,205,224 B1. 2001.

[13] Merino-Martinez, R., Vieira, A., Snellen, M., and Simons, D., "Sound quality metrics applied to aircraft components under operational conditions using a microphone array," *25^{th} AIAA/CEAS Aeroacoustics Conference, May 20 – 24 2019. Delft, The Netherlands*, , No. AIAA paper 2019–XXXX, 2019.

[14] Merino-Martinez, R., Snellen, M., and Simons, D. G., "Functional beamforming applied to imaging of flyover noise on landing aircraft," *Journal of Aircraft*, Vol. 53, No. 6, 2016, pp. 1830–1843. doi:10.2514/1.C033691, URL `http://arc.aiaa.org/doi/abs/10.2514/1.C033691`.

[15] Merino-Martinez, R., Snellen, M., and Simons, D. G., "Determination of Aircraft Noise Variability Using an Acoustic Camera," *23^{rd} International Congress on Sound and Vibration, July 10 – 14 2016, Athens, Greece*, International Inst. of Acoustics and Vibration (IIAV), Auburn, Alabama, USA., 2016.

[16] Snellen, M., Merino-Martinez, R., and Simons, D. G., "Assessment of noise level variability on landing aircraft using a phased microphone array," *Journal of Aircraft*, Vol. 54, No. 6, 2017, pp. 2173–2183. doi:10.2514/1.C033950, URL `http://dx.doi.org/10.2514/1.C033950`.

[17] Merino-Martinez, R., Heblij, S. J., Bergmans, D. H. T., Snellen, M., and Simons, D. G., "Improving Aircraft Noise Predictions by Considering the Fan Rotational Speed," *Journal of Aircraft*, 2018. doi:10.2514/1.C034849, Accepted for publication. DOI: 10.2514/1.C034849.

[18] Schluter, S., and Becker, S., "Determination of aircraft engine speed based on acoustic measurements," 2016.

[19] Vieira, A., Snellen, M., and Simons, D. G., "Assessing the shielding of engine noise by the wings for current aircraft using model predictions and measurements," *Journal of the Acoustical Society of America*, Vol. 143, No. 1, 2018, pp. 388–398. doi:10.1121/1.5020798, URL `https://asa.scitation.org/doi/full/10.1121/1.5020798`.

[20] Li, Y., Wang, X., and Zhang, D., "Control strategies for aircraft airframe noise reduction," *Chinese Journal of Aeronautics*, Vol. 26, No. 2, 2013, pp. 249–260. doi:10.1016/j.cja.2013.02.001.

[21] Merino-Martinez, R., Bertsch, L., Snellen, M., and Simons, D. G., "Analysis of landing gear noise during approach," *22^{nd} AIAA/CEAS Aeroacoustics Conference. May 30 – June 1 2016. Lyon, France*, 2016. doi:10.2514/6.2016-2769, URL `http://arc.aiaa.org/doi/pdf/10.2514/6.2016-2769`, AIAA paper 2016–2769.

[22] Hoed, W., Christensen, L., and Walker, R., "Subjective assessment of audio quality-the means and methods in the EBU," *EBU Technical Review*, , No. 40, 1997.