Objective quantification of perceived differences between measured and synthesized aircraft sounds

Sahai, Abhishek; Snellen, Mirjam; Simons, Dick

DOI
10.1016/j.ast.2017.10.035

Publication date
2018

Document Version
Accepted author manuscript

Published in
Aerospace Science and Technology

Citation (APA)
Sahai, A., Snellen, M., & Simons, D. (2018). Objective quantification of perceived differences between measured and synthesized aircraft sounds. *Aerospace Science and Technology*, *72*, 25–35. https://doi.org/10.1016/j.ast.2017.10.035

Objective quantification of perceived differences between measured and synthesized aircraft sounds

Abhishek K. Sahai[§††], Mirjam Snellen[§] and Dick G. Simons[§]

Abstract

This paper presents an approach with which perceived audible differences in aircraft sounds can be quantified and presented in an objective manner. The objective quantification of the subjectively heard audible differences is intended to serve two primary goals. It can firstly enable developers of auralization technology to make the auralized sounds more realistic by identifying in which aspects the synthesized sounds differ from their real-life counterparts and to what extent. The quantification can secondly provide an improved and more detailed means of distinguishing between aircraft sounds in general, beyond the conventional metrics of A-weighted Sound Pressure level (dBA) or Effective Perceived Noise Level (EPNL) used currently to assess aircraft noise. In this study sound quality metrics are used to quantify the differences in aircraft sounds. These metrics are widely used in other industries such as the automotive sector. Audio files of a reference aircraft, made over identical flight paths at a noise monitoring station in the vicinity of Schiphol airport, are compared in terms of both conventional and sound quality metrics for four measured and four auralized audio files. It is observed from the comparison that differences that may appear small in the conventional metrics can be significant in terms of the sound quality metrics. Significant differences in measured and synthesized sounds are observed for the aircraft considered in this study with regards to the tonal content and fluctuations in amplitude that occur over time. The conventional metrics are seen to capture the overall loudness aspect of aircraft sounds, but give no clear information regarding which spectral or temporal characteristics cause the sounds to be perceived as audibly different.

Keywords: Aircraft noise auralization; aircraft noise synthesis; aircraft sound quality; aircraft noise annoyance; acoustic perception-influenced design

1. Introduction

Aircraft noise is a very complex noise source, containing both broadband and tonal frequency components which span a wide range of frequencies. Additionally, aircraft noise can contain strong and rapid fluctuations in amplitude over time, which further add to its complexity. The noise produced by aircraft has traditionally been expressed in a specific overall value such as the maximum A-weighted Sound Pressure level (dBA), Sound Exposure Level (SEL) and also the Effective Perceived Noise Level (EPNL). Any differences in the noise two aircraft produce have therefore till now been expressed in these commonly used metrics by the aerospace community. These metrics quantify differences in overall noise impact. However, they do not indicate in which way the sounds differ from each other and which spectral or temporal characteristics are the cause of the overall difference. Further deficiencies become apparent when two aircraft have very similar noise impact values when expressed in A-weighted Sound Pressure levels and EPNL values, yet their sounds are noticeably different when heard by observers. Previous studies such as those by Hellman and Zwicker [1] and Scharf and Hellman [2] have shown that traditional metrics such as dBA and others that use it as a basis, do not suffice in clearly distinguishing between sound signatures and studies by Angerer et al. [3] have shown that they also do not correspond well to the actual perceived annoyance caused by aircraft noise, something which is an overall goal of any aircraft noise metric. Similar deficiencies have

---

[§] Aircraft Noise and Climate Effects (ANCE) Research Group, Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, 2625HS Delft, The Netherlands.
[††] *Author email:* a.k.sahai@tudelft.nl

been identified to some extent for the EPNL metric, particularly when trying to capture differences in tonal content of aircraft noise, as shown by More et al. in [4] and by Sahai and Stumpf in [5] and [6]. These conventional metrics can therefore lack in capturing important differences in aircraft sounds, such as the prominence of tonal noise in relation to broadband noise, fluctuations in amplitude over time or the ratio of high to low frequency noise for instance. The ability to objectively capture differences between aircraft sounds is of fundamental importance for reducing the adverse effects due to aircraft noise experienced by residents. Any noise reductions achieved solely by focusing on metrics that do not fully capture the individual aspects of aircraft noise may not result in a reduction of the actual annoyance experienced by residents.

The current paper aims to build on the work of Arntzen et al. [7], where a comparison of measured and synthesized aircraft noise of the same aircraft was presented and the differences in noise impact were expressed in conventional dBA and SEL metrics. It was observed that although the differences in L_{Amax} values were 0-3 dBA and in SEL values 0-4 dBA, the measured and synthesized sounds were audibly quite clearly different. Some of the audible differences observed were that the synthesized tones were too prominent, the low frequency noise had been underpredicted and that there was significantly more turbulence in the recordings due to the presence of wind-gusts. These differences were not captured by the A-weighted metrics and the results highlighted the need to express differences in aircraft sounds in an improved, objective manner. Another study, focused on the auralized sounds of future Counter Rotating Open Rotor (CROR) engines by Rizzi et al. [8], has also shown that audible changes in the quality of the sounds due to improved blade designs are not always clearly expressed in terms of EPNL values. By starting with a focus on comparison of synthetic and measured aircraft sounds, various factors can be investigated which play an important role in distinguishing between aircraft sound signatures in general. The objective distinction of aircraft sound signatures and their individual characteristics serves the second, more overall goal of designing aircraft for optimal sound, whereby aircraft designs can be optimized to meet required, more acceptable target sound signatures. This approach is intended to shift the focus from ‘low-noise’ aircraft design, which is the current practice, towards low-annoyance or ‘perception-influenced’ aircraft design. This is an approach which focuses on influencing aircraft designs towards achieving a more favorable human response to aircraft noise, as presented by Rizzi in [9] and by Diez et al. in [10] and [11].

The paper is divided into five main sections. The aircraft noise measurement and auralization methodology as explained in [7] is briefly recapitulated in Section 2. The sound assessment methodology, which is done via a combination of an Audio Assessment Module (AAM) currently being developed for automated aircraft noise audio assessment and the PULSE Reflex software of Bruel and Kjaer, is explained in Section 3. Section 3.1 explains the steps used for the implementation of the sound quality metrics of stationary loudness and sharpness in the AAM, as well as a brief background on the tonality, roughness and fluctuation strength metrics as used in PULSE Reflex. The comparison of the aircraft noise assessment in conventional and sound quality metrics is performed in Section 4 and the conclusions of the current work are presented in Section 5.

2. Noise measurement and synthesis approach

The reference aircraft used for making the comparison of measured and synthesized (i.e. auralized) aircraft noise is the Boeing 747-400 equipped with four CF6-80C2 engines. The comparison has been made for four takeoff flight paths of the 747-400, with the noise measured at a noise monitoring location near Schiphol airport Amsterdam, situated 3.8 km in front and 400 m to the right of the runway in the aircraft takeoff direction. The noise measurement has been performed using the Noise Monitoring System (NOMOS)[‡‡], which continuously measures noise from aircraft at 31 ground locations spread over the extended vicinity of Schiphol airport. Each monitoring station, including the station selected for the comparison in this study, makes use of type 1 sound meters manufactured by

---

[‡‡] NOMOS live noise monitoring available at http://noiselab.casper.aero/ams/#page=home, accessed on 24-05-2017.

Bruel and Kjaer, having a measurement accuracy of 0.7-0.9 dBA. The measurement system is coupled with radar to get information on the type of aircraft and satisfies ISO 20906 requirements for monitoring aircraft sound in airport vicinities [12]. The selected noise monitoring station was located on a grassy field, with the microphone placed on a pole at a height of 10 m in order to minimize the effect of ground reflections and absorption from the ground surface [13]. The measured audio data was provided by Schiphol airport for the initial analysis of Arntzen et al. in [7] and is considered highly accurate and reliable data, also used to make policymaking decisions and for community outreach for noise affected communities. The noise has been both measured and auralized at the same monitoring location. As can be seen in the measured audio spectrograms in Figs. 1-4 shown at the end of this section, relatively strong wind-gusts were present during the day of the measurements, which are observed as vertical spikes in the measured spectrograms. The microphone has a small wind cap but the equipment does not include a wind speed meter, which would aid in the quantification of the local wind and the resulting wind-induced turbulence. The wind cap was in this case not sufficient to shield the microphone from the strong wind-gusts and a larger wind shield would be required to minimize the effects of the gusts in the future. The first three flight paths, Flight Path (FP) 1 to 3, follow a similar trajectory with the aircraft flying closer to the monitoring station than during FP 4, where it flew further away from the monitoring station in a lateral direction. The intensities of both measured and synthesized aircraft noise are therefore higher for FP 1-3 than for FP 4, due to the shorter distance between the aircraft and observer for the first three flight paths.

The aircraft noise synthesis firstly requires the simulation of aircraft noise at the source. For this purpose, the inputs required by the fan, jet and airframe source noise prediction models were simulated over the measured flight paths. The source noise models used for the synthesis are based on NASA's Aircraft Noise Prediction Program (ANOPP), which includes the model of Heidmann [14] for fan and compressor noise, Stone [15] for jet noise and Fink [16] for airframe noise. For calculating the engine noise, which is the dominant noise source during takeoff, the engine state over the flight path was simulated using the NLR and TU Delft Gas-turbine Simulation Program (GSP) software [17], by creating a representative CF6-80C2 engine model. All the thermodynamic inputs required for the engine noise calculation were then extracted from GSP for the thrust required for the given aircraft takeoff weight, lift produced and drag experienced during the takeoff phase using relevant lift-drag polars. Since combustor and turbine noise are not dominant during takeoff, their simulations were left out of the prediction and subsequent analysis. The source noise models used in this study are semi-empirical in nature and although they do not provide a hundred percent match to measured data in terms of the spectra and directivities, particularly for the more modern aircraft of today, they are still regarded as state of the art prediction models. Their generic noise prediction capability can be applied to any conventional aircraft and engine, and their computational efficiency also makes their use highly desirable. Some audible differences between the synthesis and measurements are therefore expected. The goal here is to quantify them in an improved way to the conventionally used metrics.

Using the predicted spectral and directional information at the source, the source noise is then synthesized. Due to their very different nature, different approaches are followed for the synthesis of broadband noise and tonal noise. For auralizing tonal noise, an *additive synthesis* technique has been used, such as that used by Allen et al. in [18] and by Sahai et al. in [19] and [20], which is shown through Eqs. (1) and (2).

$$s_i(t) = A_i cos(\varphi_i(t) + \varphi_0) \tag{1}$$

$$\varphi_i(t) = 2\pi \int_{-\infty}^{t} f_i(\tau) d\tau \tag{2}$$

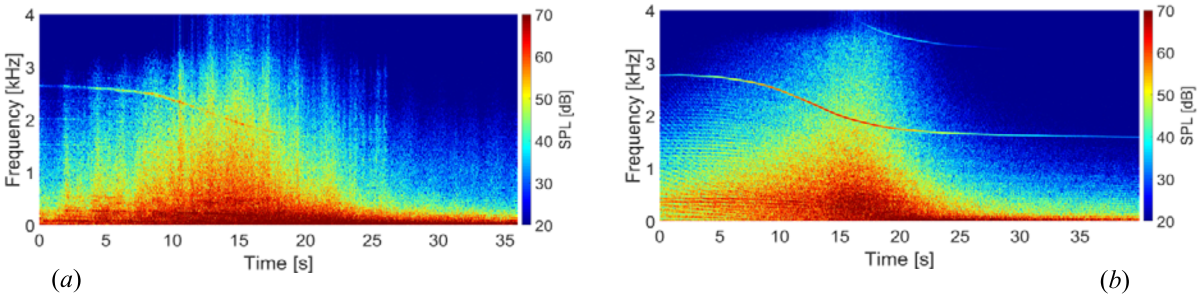
Here each tone's signal is constructed as a cosine wave having amplitude A_i, instantaneous phase φ_i and an initial phase φ_0, which is set here as a random phase offset to produce more realistic and less coherent sound. The instantaneous phase φ_i is calculated from the instantaneous frequency f_i of each tone, as shown in Eq. (2). By constructing the individual tones using this technique, the total tonal signal can be constructed by adding together all the individual tones. For the current study, the rotor-stator interaction tones and buzz-saw tones from the fan have been auralized. The fan rotor-stator interaction tones and their harmonics occur at the Blade Passage Frequency (BPF) and at its integer multiples, while the buzz-saw tones, which are shock waves produced due to the fan blade tips having local Mach numbers greater than 1, occur at multiples of the engine low pressure shaft speed (*N1*). The method of Heidmann specifies the individual magnitude of each interaction tone at the frequency it occurs, but the magnitudes of the buzz-saw tones are provided for 1/3-octave frequency bands. For this reason, by looking at in which 1/3-octave band the buzz-saw tones occur, the energy specified by Heidmann's method is divided evenly over all the buzz-saw tones in that band.

For broadband noise synthesis, an *Overlap Add* (OLA) technique has been used, which makes use of white noise as a basis to generate synthetic broadband noise. The first step in this process is to convert the 1/3-octave source noise spectra provided by the models of Stone for jet noise and Fink for airframe noise to a narrowband noise spectrum. White noise is then generated in the frequency domain and subsequently convolved with the narrowband spectra. Via an Inverse Fast Fourier Transform (IFFT), the frequency domain results are transformed to the time domain. As the aircraft flies past the measurement point, the directivity angle between the aircraft and the observer changes continuously, which results in the noise signature at the observer also changing continuously. In addition, the noise reaching the observer can change due to a change in the aircraft's thrust setting and/or high-lift device setting. Both these factors can result in audible artifacts in the synthesized sound as the aircraft flies by. In order to avoid such artifacts, the time domain signals or 'grains' per block of time are windowed using a Hann window and added with an overlap, in so-called 'hops', which have a size of around 10 percent of the block size within each time block of samples considered. This technique is referred to as the OLA technique.

These two steps provide the complete tonal and broadband noise synthesis at the source. In order to reproduce the noise impact at the noise monitoring point on the ground, the propagation and flight effects are subsequently applied to the source noise as gains and filters. The propagation effects of spherical spreading i.e. decrease in level with distance traveled from the source, atmospheric absorption according to ISO-9613-1:1993 [21] and ground reflection according to Chien-Soroka theory [22] using Delany and Bazley's [23] ground impedance model are applied to propagate the source noise signal. The atmospheric absorption attenuates higher frequencies such as from the fan more effectively than lower frequencies from the jet or airframe. The goal in this paper is to compare the predicted noise with measurements. As the microphone was placed at a height of 10 m for performing the measurements, the effects of ground reflection are in this case minimal. To account for the moving source effect, the Doppler shift is applied to the signal via a Variable Delay Line (VDL), using the time-varying time-delay between the noise emission time at the aircraft and the noise reception time at the measurement point [24]. If the time-delay results in a retarded emission time that lies in between two reception time values, the VDL performs a spline interpolation in order to avoid any resulting aliasing effects.

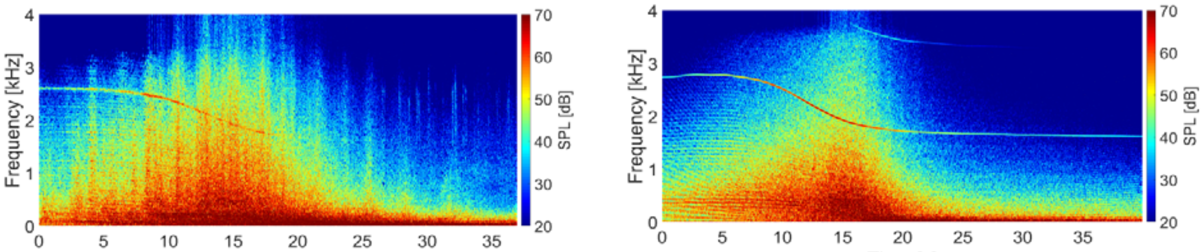
Figures 1 to 4 show the spectrograms of measured and auralized aircraft noise for the four flight paths. The measured spectrograms are shown in these figures at the left and the synthesized spectrograms are shown at the right. The vertical lines or 'spikes' mentioned earlier due to turbulent wind-gusts can be seen in all the measured audio spectrograms shown in Figs. 1-4 (a). It is evident by looking at the synthesized spectrograms in Figs. 1-4 (b) that the synthesized audio is much cleaner and lacks the turbulence that can be noticed in the measured audio's spectrograms. The wind-gusts are seen to be frequency independent and span a broadband frequency range of 0-3500 Hz. The intensity of the wind-gusts is seen to be variable and visual inspection of the vertical spikes in the measured spectrograms shows that they modulate the measured aircraft noise amplitude by 5-10 dB. Further research is needed to fully quantify the change in measured noise intensity caused by wind-gusts and develop a methodology to introduce them with the required intensity into the synthesis process. Some work in this regard is being carried out by other organizations such as in [25]. It would also be beneficial for future comparisons to limit the effect of wind-gusts by either choosing a calmer day (average wind speeds on the day of the measurement were around 3 m/s), adding wind-screens around the microphone or reducing the height of the microphone from its current position at 10 m altitude.

The fundamental fan interaction tone, beginning at around 2800 Hz, is also seen to be much cleaner and more pronounced in the synthesized audio spectrograms than what is seen in the measured audio spectrograms. This is observed for all flight paths except for FP 4, where a better match in the tonal prominence is observed. The prominence of the fundamental fan tone for FP 1-3 is in this case not only due to the fan tonal noise intensity being overpredicted in the source noise model, but also due to the fact that less broadband noise surrounds the predicted fan tones. It can also be observed that the buzz-saw tones are more pronounced and have stronger intensities in the synthesis (seen as the numerous horizontal lines in the first half of the synthesized spectrograms in Figs. 1-4). Another difference in the spectrograms that can be noticed visually is that the measured audio has on the whole higher intensity low frequency noise than what is observed in the auralized spectrograms. This is partly due to low frequency noise being underpredicted in the forward directions in the synthesis but also due to the presence of more low-frequency background noise in the measured audio. It can be mentioned here that the measured audio from the noise monitoring station is low-pass filtered to only include frequencies till 3500 Hz, as seen in the measured audio spectrograms. This was done in order to minimize data-storage owing to the high volume of air-traffic. The synthesized audio has therefore also been shown only till a frequency of 3500 Hz.



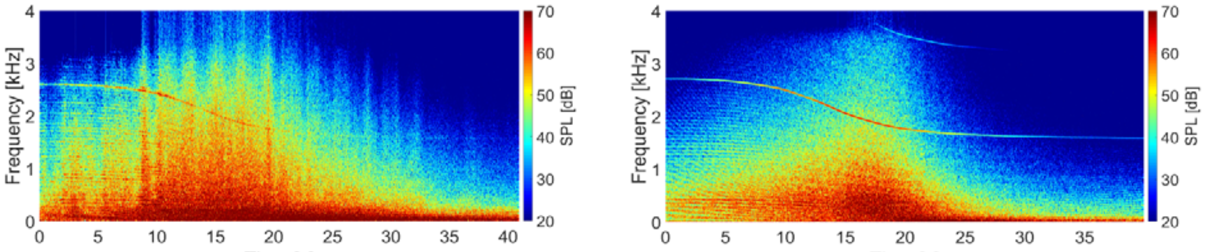
(*a*) (*b*)

Figure 1: Spectrograms at monitoring point for microphone height of 10 m for flight path 1 – (a) measured and (b) auralized



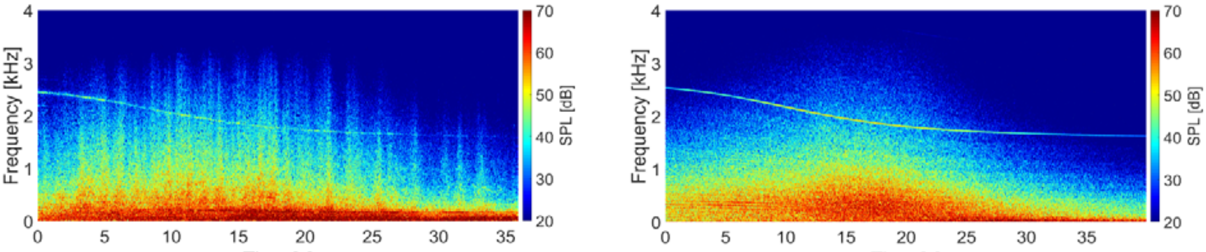
(*a*) (*b*)

Figure 2: Spectrograms at monitoring point for microphone height of 10 m for flight path 2 – (a) measured and (b) auralized



(*a*) (*b*)

Figure 3: Spectrograms at monitoring point for microphone height of 10 m for flight path 3 – (a) measured and (b) auralized



(*a*) (*b*)

Figure 4: Spectrograms at monitoring point for microphone height of 10 m for flight path 4 – (a) measured and (b) auralized

3. Noise assessment methodology

The methodology with which the noise assessment for the comparison of measured and synthesized audio is performed is explained in this section. As mentioned earlier, the noise assessment has been partly performed using an Audio Assessment Module (AAM) currently being developed at the ANCE group of TU Delft and partly with the PULSE Reflex software of Bruel and Kjaer. The AAM can currently assess any aircraft noise audio file for its A-weighted OASPL, SEL and EPNL in conventional metrics. The EPNL metric is calculated according to the procedure outlined by the Federal Aviation Administration (FAA) of the United States [26], which requires firstly the calculation of the Perceived Noise Level (PNL). The PNL is an annoyance based metric and makes use of equal noisiness curves, determined from psychoacoustic tests on a set of listeners. As the presence of discrete tones in otherwise broadband sounds makes the sound perceptually more annoying [27], a tone correction is added to the calculated PNL value. The Tone-corrected Perceived Noise Level (PNLT) metric therefore adds a tonal penalty to the PNL value if the presence of a strong protruding tone in the 1/3-octave spectrum is detected. To account for the additional effect of duration on the perceived noise level, a duration correction is added to the maximum PNLT value, PNLT_{max}, as the final step in the EPNL calculation. This yields a single value measure for the annoyance caused by aircraft noise for a single event in the unit of EPNdB.

As mentioned earlier, the A-weighting based metrics have been found to poorly capture differences in aircraft noise in previous studies by Hellamn et al. in [1] as well as [28], and also correlate poorly with perceived annoyance due to aircraft noise, as shown by Angerer et al. in [3] and More et al. in [29] as well as [4]. The studies performed in [3] as well as [4] made use of sound quality (SQ) metrics to quantify the influence of individual aircraft noise characteristics on the perceived annoyance, and compared the predicted results with observed annoyance values from listening tests. The use of SQ metrics was regarded as the most suitable approach, as it focuses on elementary perceptual features of sound for explaining and predicting how acceptable a sound is perceived to be. Previous research by the authors of this paper has also identified similar deficiencies for the EPNL metric, with regards to capturing differences in tonal content of aircraft noise in [5] and [6] as well as recently in [30]. With the aim to improve on the currently used conventional metrics, the AAM can also assess audio in sound quality metrics, in terms of stationary loudness according to the method of Zwicker outlined in DIN45631 and ISO532B [31] and time-varying loudness according to [32], and sharpness according to the methods of von Bismarck [33] and Fastl [34]. The measured and synthesized audio have been assessed using PULSE Reflex in the remaining SQ metrics of roughness and fluctuation strength according to the methods of Zwicker and Fastl [34], and tonality according to Terhardt's method [35]. It is intended in the future to extend the capability of the AAM to these remaining SQ metrics as well. The overall goal of developing the AAM is to have an automated aircraft audio assessment capability that can be integrated in an aircraft design chain for performing perception-influenced aircraft design studies, as has been proposed for the design of products from other industries by Davies in [36] and Lyon in [37]. For this purpose, as the SQ metrics relate strongly to the perceived psychoacoustic annoyance, either the individual SQ metrics or a chosen combination of these metrics in the form of an overall 'perceived annoyance' metric could be used. Such models have been suggested by Zwicker for generic product sounds [34] and by More et al. specifically for aircraft noise, by adding a weighted contribution of tonality to the predicted psychoacoustic annoyance [4], besides the weighted contributions of loudness, roughness, fluctuation strength and sharpness. Finding a suitable metric that accurately predicts the perceived annoyance due to aircraft noise remains an active field of research as seen in the studies of More in [4], Sahai in [6] and in the research projects SEFA [38] and COSMA [39] sponsored by the European Commission.

3.1 Assessment in sound quality metrics

The methodology with which the stationary loudness and sharpness metrics were implemented in the AAM is briefly explained in this sub-section. Furthermore, a brief background to the SQ metrics of tonality, roughness and fluctuation strength is also provided, to allow a better understanding of what the metrics convey with regards to the observed differences in Section 4.

3.1.1 *Stationary loudness:* Zwicker's stationary loudness calculation according to DIN45631 and ISO532B in the AAM uses as inputs 1/3-octave spectra obtained by applying a 1/3-octave filter to the pressure time signal. The stationary loudness calculation involves three primary steps to determine the loudness perception:

i. Modeling the transmission of the acoustic signal through the outer and middle ear;
ii. Estimation of an unmasked specific loudness pattern by computing a main specific loudness pattern, and then applying spectral masking curves to determine the loudness that is unmasked;
iii. Integration of the unmasked specific loudness pattern to get an overall loudness value in sone.

In order to model the transmission characteristics of the human ear as well as the spectral masking effects, Zwicker's method divides the audible frequency range into a series of frequency bands called critical bands, *z*, given the unit of Bark. The critical bandwidth for each critical band corresponds to the frequency resolution of the human ear. A sound component located in one critical band produces an excitation pattern that spreads over several critical bands. The level of the sound in each critical band is converted to an excitation level $L_E(z)$, which spreads over several adjacent critical bands. By checking if the excitation level in each critical band lies above the level of the threshold of hearing $L_{TQ}(z)$ in that critical band, the *main specific loudness* N'_{main}in sone/Bark, according to Zwicker [31] is calculated as:

$$N'_{main}(z) = 0.0635 \cdot 10^{0.025 L_{TQ}(z)} \left[\left(0.75 + 0.25 \cdot 10^{0.1(L_E(z) - L_{TQ}(z))} \right)^{0.25} - 1 \right] \tag{3}$$

A check is then made to see if the excitation in each critical band is being masked by the excitation pattern from a component located in another critical band, a so-called *accessory loudness*. By accounting for the masking in each critical band, an unmasked specific loudness value $N'_{unmasked}$ in each critical band is determined. The unmasked specific loudness pattern is then integrated over all the 24 critical bands to determine the total loudness value *N* in sone [34]:

$$N = \int_0^{24} N'_{unmasked}(z) dz \tag{4}$$

3.1.2 *Sharpness:* Both the methods of von Bismarck and Fastl implemented in the AAM determine the sharpness of a sound by making use of a weighted first moment of the unmasked specific loudness $N'_{unmasked}$. Von Bismarck's calculation uses a weighting function *g*(*z*), to weigh all spectral content at and above 16 Bark (i.e. from 2700 Hz) increasingly more heavily. The higher the frequency of a component, the higher will be its contribution therefore in the sharpness calculation according to von Bismarck [33]:

$$g(z) = \begin{Bmatrix} 1 & z \leq 16 \\ 0.066 e^{0.171 z} & z > 16 \end{Bmatrix} \tag{5}$$

The overall sharpness value *S* in the unit of acum is then calculated using Eq. (6) using von Bismarck's method [33], where the proportionality constant *c* has a value of 0.11 and *N* is the total stationary loudness computed using Eq. (4) for the considered 1/3-octave spectrum.

$$S = c \left[\frac{\int_0^{24} g(z)\, N'_{unmasked}(z)\, z\, dz}{N} \right] \tag{6}$$

The method of Fastl uses a slightly different sharpness weighting function, also determined from psychoacoustic surveys with test audiences [34]. This weighting function however also gives very similar sharpness values to those obtained using von Bismarck's method, when applied to aircraft noise.

The remaining SQ metrics of Terhardt's tonality and Zwicker and Fastl's roughness and fluctuation strength have been applied using PULSE Reflex. A detailed and exhaustive description of these metrics and how they could be implemented in the AAM would be too lengthy for the goals of the current paper. A brief explanation of the metrics and how they are calculated is nonetheless provided, in order to clarify what the metrics essentially measure and what meaning they convey.

3.1.3 *Tonality:* Tonality is a measure of the perceived strength of unmasked tonal energy present within a complex sound. Although the tonality metric as calculated according to the method of Aures [24] is recognized as a sound quality metric, the metric has not yet been standardized. Aures' tonality metric is partly based on Terhardt's tonality metric, which has been applied in PULSE Reflex. Both methods begin by computing an SPL excess of tonal components in a sound, to determine which tones are aurally relevant for tonality perception. Both methods take into account masking of tones as well as the dependence of tonality on the excess level and frequency of the tones. The primary difference between both metrics is that Terhardt's metric focuses on the spectral pitches of all tonal components and then weighs them to determine an overall virtual pitch of the sound. Aures' tonality metric determines the overall tonal perception and euphony, or the 'tonality' of a sound, by considering additionally the influence of bandwidth of tonal components, as well as their loudness relative to the overall loudness.

3.1.4 *Roughness:* Roughness is the sensation produced by sounds containing fast loudness fluctuations in the order of 15-300 cycles per second, with maximum roughness sensation experienced at around 60-70 cycles per second. The roughness metric has the aim of quantifying how the rapid amplitude modulations of a sound are perceived by the human ear. The metric is also not yet fully standardized and a number of methods exist for roughness calculation, with most using a relation of the form according to Zwicker and Fastl [34]:

$$R = 0.3\, f_{mod} \int_0^{24} \Delta L(z)dz \tag{7}$$

Equation (7) gives the roughness R in the unit of asper, with f_{mod} the modulation frequency of the sound in kHz and ΔL the temporal masking depth, or the 'subjective duration'. The temporal masking depth can be defined as the difference between the actual and perceived modulation depths. High modulation frequencies reduce the perceived modulation depth, making it differ from the actual modulation depth. The temporal masking depth has been found to be a function of frequency and therefore has to be integrated over the 24 critical bands for the roughness calculation. The method of Zwicker and Fastl proposes determining the corresponding variations of time-varying specific loudness to estimate ΔL. The influence of buzz-saw tones may be captured by the roughness metric in this paper.

3.1.5 *Fluctuation strength:* Fluctuation strength quantifies the subjective perception of slow fluctuations in loudness of the order of 1-16 cycles per second. For faster modulations above around 20 Hz, the perception of roughness begins to dominate over fluctuation strength. The fluctuation strength in units of vacil, according to the method of Zwicker and Fastl, can be calculated using Eq. (8).

$$FS = \frac{0.008 \int_0^{24} \Delta L(z)dz}{\left(\frac{f_{mod}}{4}\right)+\left(\frac{4}{f_{mod}}\right)} \tag{8}$$

As maximal fluctuation strength is experienced by the typical listener for a modulation frequency of around 4 Hz, the calculation of the fluctuation strength takes this into account by providing a term with a ratio of the actual modulation frequency (in Hz) to a modulation frequency of 4 Hz. ΔL is here again the temporal masking depth, which is replaced with the variations in time-varying specific loudness, similar to the roughness metric. For the purposes of this paper, the fluctuation strength metric may help in distinguishing between the various aircraft sounds if any slow modulations in their amplitude over time are present.

4. Objective audio comparison in conventional and sound quality metrics

A combination of the AAM and the PULSE Reflex software was used to compare the measured and synthesized audio files for each of the four flight paths. This comparison has been split into the metric values for the conventional L_{Amax}, SEL and EPNL metrics in Section 4.1 and in the SQ metrics in Section 4.2. A preliminary comparison of the flight paths was presented by the authors in [40].

4.1 Comparison in conventional aircraft noise metrics

Table 1: Comparison of measured and synthesized aircraft sounds in conventional L_{Amax}, SEL and EPNL metrics

Flight Path	Measured L_{Amax} [dBA]	Synthesized L_{Amax} [dBA]	Δ [dBA]	Measured SEL [dBA]	Synthesized SEL [dBA]	Δ [dBA]	Measured EPNL [EPNdB]	Synthesized EPNL [EPNdB]	Δ [EPNdB]
1	82.0	83.1	+1.1	90.8	90.8	0.0	93.4	92.8	-0.6
2	83.7	84.0	+0.3	93.0	91.5	-1.5	95.3	93.7	-1.6
3	83.3	81.8	-1.5	94.1	90.5	-3.6	95.9	92.3	-3.6
4	74.0	75.1	+1.1	85.0	86.0	+1.0	88.5	87.2	-1.3

It can be seen in Table 1 that the differences between the measured and synthesized sounds for each of the flight paths in the L_{Amax} values are quite low, ranging from 0.3 to 1.5 dBA. The differences for FP 1, 2 and 4 in terms of the SEL and EPNL metrics are also of the same order, between 0 and 1.5 dBA SEL and 0.6 and 1.6 EPNdB. For FP 3 however, the differences in terms of these metrics are much higher, at 3.6 dBA SEL and 3.6 EPNdB, which indicates that there might be quite significant differences between these sounds. The SEL and EPNL metrics indicate that there is a difference between the sounds, but they do not indicate in which way these sounds are actually different and which individual sound characteristics or components are causing the audible differences. Fig. 5 shows the OASPL vs time history for FP 1 and FP 3 to illustrate the L_A vs. time variation, from which it can be seen that the synthesized and measured L_{Amax} values are quite close for both of these flight paths. For FP 3 however, large differences in OASPL can be seen before and after the aircraft is at its closest point to the noise monitoring station. This difference is indicated to some extent by the SEL values in Table 1 but it remains unclear what the primary cause could be of this larger SEL difference and in which way the synthesized sound could be modified to reduce the observed difference in SEL values. The differences in the EPNL metric are similar to those for the SEL metric for all flight paths except for FP4. Similar comments can be made for the differences in EPNL values, regarding lack of information about the causes of the differences and how they could be reduced or minimized.

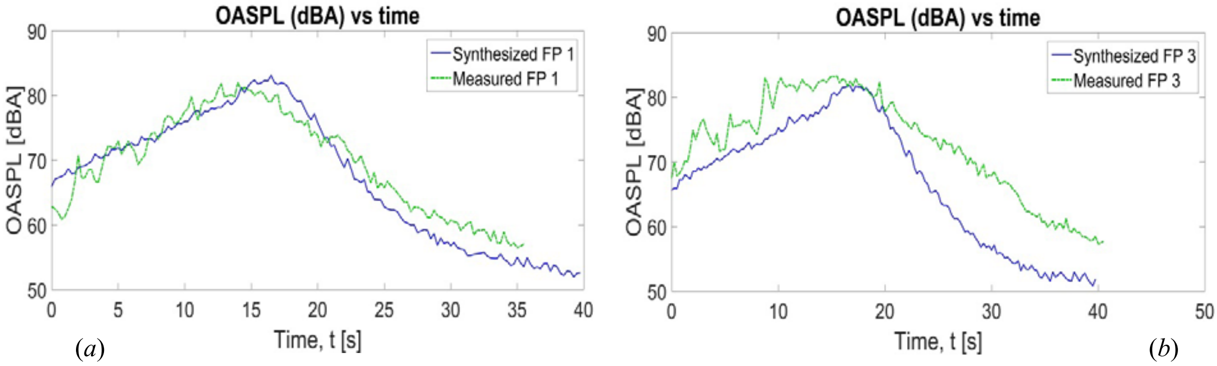


Figure 5: OASPL (dBA) vs time comparison of measured and synthesized sounds – (a) FP 1, ΔL_{Amax} = 1.1 dBA, (b) FP 3, ΔL_{Amax} = -1.5 dBA

It is for this reason that the analysis is extended to the SQ metrics, which focus on individual sound characteristics and can indicate more clearly how these sounds differ from each other and how they could be potentially modified to result in a closer match. As the conventional metrics evidently do not capture several aircraft noise characteristics, some of which are known contributors to the perceived annoyance, optimizing aircraft sounds in these metrics may ultimately also not result in a reduced annoyance experienced by residents.

4.2 Comparison in sound quality metrics

4.2.1 *Loudness comparison*: Table 2 shows the sound comparison in terms of the stationary loudness and time-varying loudness metrics computed using the AAM. The loudness values that were exceeded for 5% of the time-history, N_5, have been compared primarily because these values correspond best to the actually perceived annoyance by listeners, and the 5% excess values are also commonly used in the field of sound engineering and psychoacoustics [4]. The differences in the SQ metrics have been presented as relative differences to the measured sound values, as the SQ metrics are all linear and absolute differences in these metrics may not carry the same familiarity as they do for the conventional metrics.

Table 2: Comparison of measured and synthesized aircraft sounds in stationary and time-varying loudness metrics exceeded for 5% of the time-history

Flight Path	Measured stationary N_5 [sone]	Synthesized stationary N_5 [sone]	Δ [%]	Measured time-varying N_5 [sone]	Synthesized time-varying N_5 [sone]	Δ [%]
1	48.0	47.9	-0.2	49.5	48.6	-1.8
2	54.9	51.1	-6.9	57.2	52.1	-8.9
3	53.8	45.3	-15.8	55.6	46.4	-16.5
4	27.6	27.8	+0.7	28.7	28.4	-1.0

It can be seen in Table 2 that the differences between the measured and synthesized sounds for all flight paths in terms of loudness resemble those observed for the SEL and EPNL metrics in Table 1. The differences in both stationary and time-varying loudness are minimal for FP 1 and FP 4, but larger for FP 2 and FP 3. To show visually how the loudness metric differs from the dBA based metrics, Fig. 6 shows the variation in stationary loudness over time for FP 1 and FP 3, analogous to the OASPL variation presented in Fig. 5. It can be seen in Fig. 6 that the loudness variation for FP 1 is very similar to the OASPL variation but the differences for FP 3 are much more amplified in terms of loudness than they were in terms of OASPL. The perceived loudness of the auralized FP 3 sound is therefore clearly lower over its entire time-history than the perceived loudness of the measured FP 3 sound. The loudness metric therefore indicates that the synthesized sound would have to be made louder in order to bring it closer to the measured sound. Figure 6 shows that the synthesized loudness is lower than the measured loudness both in the forward as well as aft directions, indicating that both the fan inlet and exhaust noise, as well as possibly the jet noise are being underpredicted in the synthesis for FP3. As the fan tones appear to be overpredicted in the synthesis, the fan broadband noise will likely need to be increased at the source to reduce the differences in loudness. The influence of jet noise on the loudness in the aft directions would also need to be further investigated.

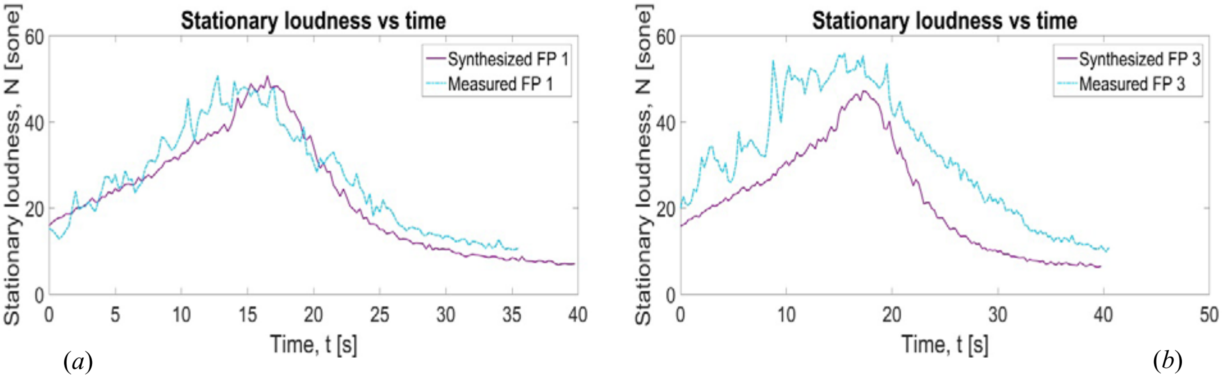


Figure 6: Stationary loudness comparison of measured and synthesized sounds – (a) FP 1, ΔN_5 = -0.1 sone (-0.2%), (b) FP 3, ΔN_5 = -8.5 sone (-15.8%)

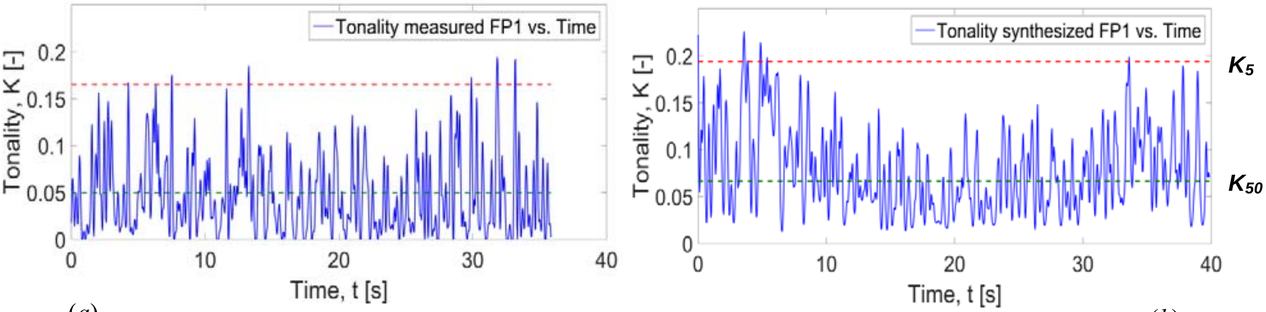
4.2.2 *Tonality comparison:* It was observed while listening to the synthesized audio files as well as from the spectrograms in Figs. 1-4 that one of the most noticeable differences between the sounds was that the tonal content is overpredicted in the synthesis. The tonality metric is intended to focus exactly on this characteristic of a sound and could identify this difference, which was not specifically captured by the conventional metrics in Table 1. Table 3 shows the comparison in terms of the tonality, K via the value exceeded for 5% of the time-history, K_5, as well as via the tonality value exceeded for 50% of the time-history, K_{50}, which is used by some standard psychoacoustics software and gives a measure of the average tonality over the entire time-history. It can be seen that the differences in tonality are significant for FP 1-3 and minimal for FP 4. Although limited references exist for the tonality of aircraft sounds, the tonality values presented in Table 3 are within the range of values for aircraft noise as shown by More in [4] and by Minard et al. in [41]. The values presented in Table 3 indicate that the tonal perception of the synthesized sounds for FP 1-3 is clearly different to the tonal perception of the corresponding measured sounds. For comparison with other more conventional tonal metrics such as the Prominence Ratio (PR) according to ECMA-74 [42], the PR_5 values for FP 1-3 are 4 dB, 11 dB and 10 dB respectively. The corresponding PR_5 values for the measured FP 1-3 sounds are: 2 dB, 3.5 dB and 6 dB respectively. The difference in tonal content is therefore quite significant for the analyzed sounds and the synthesized tones are overall much more prominent.

Fig. 7 visualizes what the tonality variation over time looks like for the measured and synthesized sounds for FP 1. The figure has been recreated using data exported from PULSE Reflex for both sounds and shows that for the measured FP 1 sound, the tonality is on the whole lower but also follows a somewhat flatter variation over the time-history. The tonality for the synthesized FP 1 sound however is seen to peak towards the start of the flyover and towards the end of the flyover. The reduction in tonality as the aircraft comes closer to the monitoring station is due to the fact that the fan tones become less prominent in relation to the surrounding broadband noise and are furthermore masked by the lower frequency jet noise, which gains in intensity as the aircraft approaches the monitoring location. The red dashed line in Fig. 7 is K_5 and the green dashed line is K_{50}.

Table 3: Comparison of measured and synthesized aircraft sounds in the tonality metric exceeded for 5% and 50% of the time-history

Flight Path	Measured K_5 [-]	Synthesized K_5 [-]	Δ [%]	Measured K_{50} [-]	Synthesized K_{50} [-]	Δ [%]
1	0.165	0.194	+17.6	0.0493	0.0690	+33.9
2	0.189	0.216	+14.3	0.0460	0.0716	+55.7
3	0.157	0.196	+24.8	0.0459	0.0665	+44.9
4	0.184	0.188	+2.2	0.0512	0.0534	+4.3

By objectively quantifying the subjectively perceived differences in the tonal content via the tonality metric, developers of auralization technology can attempt to adapt the tonal content's intensity as well as that of the surrounding broadband noise around each fan tone, in order to bring the synthesis closer to the measurements. This will ideally have to be performed at the source noise modeling stage, where the fan noise intensity being emitted from the engines, as modeled using the method of Heidmann, will have to be reduced for the tones and increased slightly for the broadband noise.



(*a*)

(*b*)

Figure 7: Tonality comparison of measured and synthesized sounds for FP 1 – (a) measured, (b) synthesized, ΔK_5 = 0.029 (+17.6%) and ΔK_{50} = 0.02 (+33.9%)

4.2.3 *Roughness and fluctuation strength comparison:* Table 4 focuses on the temporal variations in the sounds and shows the comparison in the roughness, R and fluctuation strength, FS, metrics. The roughness metric quantifies the subjective perception of fast amplitude modulations of a sound, of the order of 50-90 Hz whereas fluctuation strength measures the subjective perception of slow amplitude modulation of the order of 1-16 Hz. It can be seen in Table 4 that the measured and synthesized roughness values for FP 1 and FP 4 are quite close to each other (2.76 asper compared to 2.66 asper and 1.79 asper compared to 1.69 asper, respectively). For FP 2 and FP 3 however, the measured sounds are rougher than the synthesized sounds by a reasonable to large amount. Two factors can cause in this case the strong perception of roughness – 1) fast amplitude fluctuations due to the presence of turbulence and 2) the occurrence of buzz-saw tones. The measured sounds are primarily rougher due to the presence of turbulence. It was noticed in the spectrograms of the synthesized audio in Figs. 1-4 (right) that the buzz-saw tones had been largely overpredicted in the synthesis, which results in an increase in the synthesized roughness value. The roughness values for the synthesized sounds for FP 1 and FP 4 are therefore relatively high even though no turbulence is included in the synthesis. For FP 2 and FP 3, the increase in roughness due to the strong buzz-saw tones predicted in the synthesis is however not high enough, to match the high roughness values produced due to the presence of turbulence in the measurements. It can be concluded that the roughness metric is unable to fully distinguish between the effects of turbulence and the under- or overprediction of buzz-saw noise. Nonetheless, a difference in roughness between the measured and auralized sounds indicates that a difference in turbulence and/or buzz-saw noise is present and needs inspection. Part of this ambiguity can be resolved using the fluctuation strength metric, as is further explained below.

Table 4: Comparison of measured and synthesized aircraft sounds in roughness and fluctuation strength metrics, exceeded for 5% of the time-history

Flight Path	Measured R_5 [asper]	Synthesized R_5 [asper]	Δ [%]	Measured FS_5 [vacil]	Synthesized FS_5 [vacil]	Δ [%]
1	2.76	2.66	-3.6	1.95	1.14	-41.5
2	3.22	2.73	-15.2	2.18	1.13	-48.2
3	3.21	2.43	-24.3	1.91	1.18	-38.2
4	1.79	1.69	-5.6	1.55	0.970	-37.4

The amplitude fluctuations due to turbulence can be both fast and slow, and contribute to a large extent to the perceived differences between the sounds. The effect of the wind-gusts can therefore be reflected in both roughness and fluctuation strength, depending on whether the fluctuations in amplitude were fast or slow in nature. The presence of slow amplitude fluctuations due to the wind-gusts can be seen in Table 4 for all measured flight paths through the fluctuation strength metric, which has consistently higher values for the measured sounds. The comparison indicates that these slow amplitude fluctuations due to turbulence were missing in the auralizations and by what amount they could be introduced to make them sound more realistic. The relative differences in the measured and synthesized *FS* values that were exceeded for 5% of the time-history are between 40-50% for all four flight paths. Based on the comparison of values in Table 4, it can be said that the conventional metrics do not identify these differences in temporal characteristics and the SQ metrics, particularly the fluctuation strength metric applied to the current study, offer a much clearer picture of the perceived differences. Fig. 8 shows the *FS* over time variation for FP 1 and the clear difference the two sounds have in terms of their fluctuation strength.

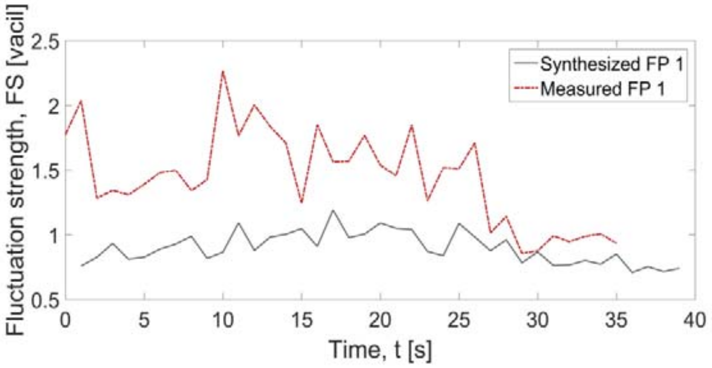


Figure 8: Fluctuation strength comparison of measured and synthesized sounds for FP 1, Δ FS_5 = -0.81 vacil (-41.5%)

4.2.4 *Sharpness comparison:* The final SQ metric used to quantify the differences between the measured and synthesized sounds is the sharpness metric, which can potentially capture differences in the high and low frequency content of the sounds. It was mentioned earlier that the measured sounds had higher low frequency content partly due to more low frequency noise emitted by the real aircraft in the forward directions but also due to the presence of more low frequency background noise. Although the background noise is not modeled and some differences in the low frequency content will therefore exist, it was intended that the sharpness metric would be able to capture these differences in frequency content. Table 5 shows however that the differences in the sharpness value exceeded for 5% of the time-history are very low, except for FP 3, where a reasonable difference between the sharpness of the measured and synthesized sound can be observed. The sharpness values in Table 5 are consistent with values presented in other references such as [43] by Rizzi et al.

Table 5: Comparison of measured and synthesized aircraft sounds in the sharpness metric exceeded for 5% of the time-history

Flight Path	Measured S_5 [acum]	Synthesized S_5 [acum]	Δ [%]
1	0.958	0.952	-0.6
2	1.05	0.998	-5.0
3	1.06	0.936	-11.7
4	0.803	0.819	+2.0

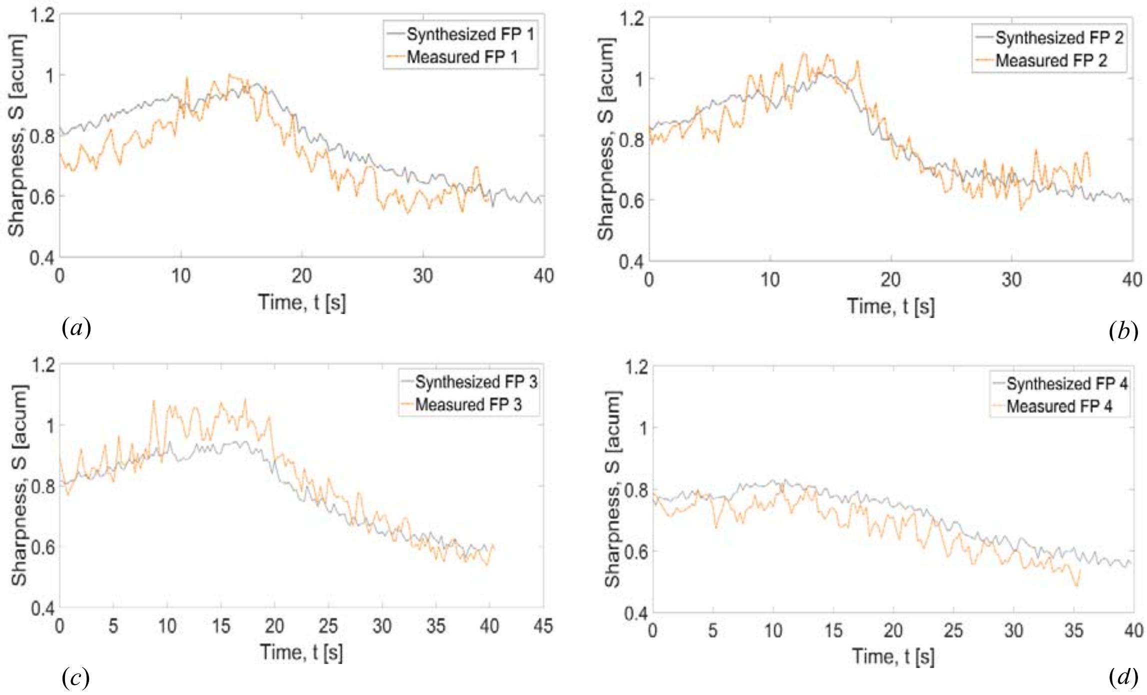


Figure 9: Sharpness vs time comparison for measured and synthesized sounds – (a) FP1, $\Delta S_5 = -0.006$ acum (-0.6%), (b) FP2, $\Delta S_5 = 0.05$ acum (-5%), (c) FP3, $\Delta S_5 = -0.124$ acum (-11.7%), (d) FP4, $\Delta S_5 = 0.016$ acum (+2%)

Looking at the sharpness variation plots for each flight path in Fig. 9, it can be seen that the synthesized and measured sharpness variations largely overlap each other. The differences for FP 3 can be seen around the time of the maximum but the values closely match each other for the remaining parts of the flyover. This indicates that the sharpness metric in its current form, whereby high frequency noise above 16 Bark or 2700 Hz is weighted increasingly heavily, is not capturing differences in frequency content that are relevant for aircraft noise purposes.

This observation is further supported by the fact that from all the SQ metrics that contribute to the perceived annoyance due to aircraft noise, sharpness variation is seen to have little to no effect [3], [4]. This is primarily due to the fact that aircraft noise signatures on the ground in general do not vary as much in terms of high frequency noise above 2700 Hz, as they do for low frequency noise. A different approach to capturing differences in high and low frequency content of aircraft noise would therefore be needed to adequately capture the perceived differences. In addition, research is needed to analyze how any new metric developed for such a purpose corresponds to the perceived annoyance.

5. Conclusions and outlook

This paper provided a means of objectively quantifying subjective differences in aircraft sounds beyond the conventional metrics used currently to assess aircraft noise. The comparison was made by focusing on measured and synthesized sounds of an aircraft during takeoff at a representative noise monitoring location. By listening to the measured and synthesized sounds, clear differences could be noticed. However, it was observed that the differences between the sounds were for the most case small when expressed in the conventional L_{Amax}, SEL and EPNL metrics and furthermore did not clearly indicate which aircraft sound characteristics were the cause of the differences. By using sound quality metrics to quantify the differences, it was observed in a clearer way which aircraft noise characteristics were different and by what objective amounts. The following differences were observed:

- It was found that the synthesized aircraft noise had overall slightly lower loudness than the measured noise, although the differences in loudness were not very large except for FP 3. The differences in terms of loudness for FP 3 were much more amplified in comparison to the OASPL metric, and showed the synthesis having a lower loudness than the measurement both in forward and aft directions. The fan broadband noise would likely need to be increased in intensity in the source noise models to reduce the differences in the forward directions. In the aft directions, both the fan exhaust noise as well as possibly the jet noise may have to be increased to reduce the difference in loudness.
- Significant differences in the tonality of the sounds were observed. The auralized sounds were perceived as clearly more tonal than the measured sounds. Reducing the fan tonal intensity and its prominence in the source noise models would aid in bringing the synthesis closer to the measurements. Reducing the prominence of the tones would also require increasing the fan broadband noise surrounding the tones.
- Significant differences were also observed in terms of temporal characteristics. The roughness metric captured these differences to some extent for FP 2 and FP 3, but could not fully distinguish between the increase in roughness due to turbulence and the increase due to overprediction of buzz-saw tones in the synthesis. The presence of turbulence in the measurements was displayed more clearly via the fluctuation strength metric. Differences in terms of fluctuation strength showed that the auralized noise had 40-50% less slow amplitude fluctuations than the measured noise. Inclusion of these temporal variations would improve the comparison between the synthesized and measured sounds and make the synthesis more realistic. Further research is however needed to quantify the changes in amplitude induced by the wind-gusts and introducing them with the required intensity in to the synthesis process.
- An analysis of the sharpness metric to test its applicability to aircraft noise showed that differences in frequency content are not captured by the sharpness metric in its current form. An alternate metric would be needed to capture the perceived differences in frequency content of aircraft noise, and also relate them to changes in perceived annoyance.

The analysis presented in this paper highlights the need to look beyond the conventional metrics used today to assess aircraft noise in order to be able to distinguish clearly between aircraft sounds. The ability to objectively quantify differences between aircraft sounds and their individual characteristics can not only allow improvement of the aircraft noise auralization capability, but also aid in modifying the synthesized sounds to specific target sounds that have a high likelihood of being judged as less annoying by the residents. The analysis in this paper has been carried out on a limited dataset of aircraft sounds. More concrete conclusions regarding the observed trends in aircraft noise sound quality, as well as how the aircraft sounds differ from each other, can only be drawn with further analysis of a larger dataset of aircraft sounds. This has been planned as future work, whereby a large dataset of aircraft sounds will be analyzed for sound quality and predicted annoyance changes, and it will be investigated how the metrics scale across different aircraft types.

References

[1] R. P. Hellman and E. Zwicker, "Why can a decrease in dB(A) produce an increase in loudness?," *Journal of the Acoustical Society of America,* vol. 82, no. 5, pp. 1700-1705, 1987.

[2] B. Scharf and R. P. Hellman, "Comparison of various methods for predicting the loudness and acceptability of noise, Part II: Effects of spectral pattern and tonal components," U.S. Environmental Protection Agency, Report 550/9-79-102, 1977.

[3] J. R. Angerer, R. A. Erickson and D. A. McCurdy, "Development of an annoyance model based upon elementary auditory sensations for steady-state aircraft interior noise containing tonal components," NASA TM 104147, 1991.

[4] S. More, "Aircraft noise metrics and characteristics," PhD thesis Purdue University written as PARTNER Project 24 Report COE-2011-004, 2011.

[5] A. K. Sahai and E. Stumpf, "Incorporating and minimizing aircraft noise annoyance during conceptual aircraft design," in *20th AIAA/CEAS Aeroacoustics Conference - Part of AVIATION forum*, Atlanta, GA, USA, 2014.

[6] A. K. Sahai, "Consideration of aircraft noise annoyance during conceptual aircraft design," PhD Dissertation, ILR, RWTH Aachen University, Aachen, Germany, 2016.

[7] M. Arntzen and D. Simons, "Modeling and sysnthesis of aircraft flyover noise," *Applied Acoustics,* vol. 84, pp. 99-106, 2014.

[8] S. A. Rizzi, D. B. Stephens , J. J. Berton, D. E. Van Zante, J. P. Wojno and T. W. Goerig, "Auralization of flyover noise from open rotor engines using model scale test data," in *20th AIAA/CEAS Aeroacoustics Conference AIAA-2014-2750*, Atlanta, GA, USA, 2014.

[9] S. A. Rizzi, "Toward reduced aircraft community noise impact via a percepton-influenced design approach," in *Internoise*, Hamburg, 2016.

[10] M. Diez, U. Iemma and V. Marchese, "A sound-matching-based approach for aircraft noise annoyance alleviation via MDO," in *13th AIAA/CEAS Aeroacoustics conference*, Rome, Italy, 2007.

[11] M. Diez and U. Iemma, " Multidisciplinary conceptual design optimization of aircraft using a sound-matching based objective function," *Engineering Optimization,* vol. 44, no. 4, pp. 591-612, 2012.

[12] "ISO 20906:2009, Acoustics - Unattended monitoring of aircraft sound in the vicinity of airports," International Organization for Standardization, 2009.

[13] G. J. Ruijgrok, Elements of aviation acoustics, VSSD, Delft University Press, 1999.

[14] M. F. Heidmann, "Interim prediction method for fan and compressor source noise," NASA TM-X-71763, 1979.

[15] J. R. Stone, "An improved prediction method for the noise generated in flight by circular jets," NASA TM 81470, 1980.

[16] M. R. Fink, "Airframe noise prediction method," FAA-RD-77-29, 1977.

[17] W. Visser and M. Broomhead, "GSP - A generic object oriented gas turbine simulation environment," NLR, 2000.

[18] M. Allen, S. A. Rizzi, R. Burdisso and S. Okcu, "Analysis and Synthesis of Tonal Aircraft Noise Sources," in *18th AIAA/CEAS Aeroacoustics Conference (33rd AIAA Aeroacoustics Conference) AIAA-2012-2078*, Colorado Springs, USA, 2012.

[19] A. K. Sahai, E. Anton and E. Stumpf, "An interdisciplinary apporach for a virtual air-traffic system simulation for subjective assessment in VR environments," in *28th International Congress of the Aeronautical Sciences (ICAS)*, Brisbane, Australia, 2012.

[20] A. K. Sahai, F. Wefers, S. Pick, E. Stumpf, M. Vorländer and T. Kuhlen, "Interactive simulation of aircraft

noise in aural and visual virtual environments," *Applied Acoustics,* vol. 101, pp. 24-38, 2016.

[21] International Organization for Standardization (ISO), "Acoustics - Attenuation of sound during propagation outdoors - Part 1: Calculation of absorption of sound by the atmosphere," ISO-9613-1:1993(E), 1993.

[22] C. F. Chien and W. W. Soroka, "Sound propagation along an impedance plane," *Journal of Sound and Vibration,* vol. 43, no. 1, pp. 9-20, 1975.

[23] M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," *Applied Acoustics,* vol. 3, no. 2, pp. 105-116, 1970.

[24] S. A. Rizzi and B. M. Sullivan, "Synthesis of Virtual Environments for Aircraft Community Noise Impact Studies," in *11th AIAA/CEAS Aeroacoustics Conference (26th AIAA Aeroacoustics Conference) 23-25 May 2005*, Monterey, California, 2005.

[25] F. Rietdijk, J. Forssen and K. Heutschi, "Generating sequences of acoustic scintillations," *Acta Acustica United With Acustica,* vol. 103, pp. 331-338, 2017.

[26] Federal Aviation Administration (FAA), "Noise standards: Aircraft type and airworthiness certification, calculation of effective perceived noise level from measured data," Federal Aviation Regulations Part 36, Appendix A2 to Part 36 - Section A36.4, 2002.

[27] K. D. Kryter and K. S. Pearsons, "Judging noisiness of a band of random noise containing an audible pure tone," *Journal of the Acoustical Society of America,* vol. 38, no. 1, pp. 106-112, 1965.

[28] R. P. Hellmann, "Contribution of tonal components to the overall loudness, annoyance and noisinessof noise: Relation between single tones and noise spectral shape," NASA CR-3892, 1985.

[29] S. More and P. Davies, "An examination of the influence of tonalness on ratings of aircraft noise," in *Proceedings of the Sound Quality Symposium 2008*, Dearborn, Michigan, USA, 2008.

[30] A. K. Sahai, M. Snellen, D. G. Simons and E. Stumpf, "Aircraft design optimization for reducing community noise exposure based on annoyance metrics," *AIAA Journal of Aircraft,* 2017.

[31] E. Zwicker, H. Fastl and C. Dallamayr, "Basic program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532B," *Acustica,* vol. 55, pp. 63-67, 1982.

[32] E. Zwicker, "Procedure for calculating the loudness of temporally variable sounds," *Journal of the Acoustical Society of America,* vol. 62, no. 3, pp. 675-682, 1977.

[33] G. von Bismarck, "Sharpness as an attribute of the timbre of sounds," *Acustica,* vol. 30, pp. 159-172, 1974.

[34] E. Zwicker and H. Fastl, Psychoacoustics - Facts and Models, Springer Verlag, 1990.

[35] E. Terhardt, G. Stoll and M. Seewan, "Algorithm for extraction of pitch and pitch salience from complex tonal signals," *Journal of the Acoustical Society of America,* vol. 71, no. 3, pp. 679-688, 1982.

[36] P. Davies, "Perception-based engineering: Integrating human responses into product and system design," *The Bridge,* vol. 37, no. 3, pp. 18-24, 2007.

[37] R. H. Lyon, " Product sound quality – from perception to design," *Sound Vibration,* vol. 37, no. 3, pp. 18-23, 2003.

[38] SEFA project consortium, "SEFA report summary," European Commission, Research and Development Information Service, Available from: http://cordis.europa.eu/result/rcn/47580_en.html, 2011.

[39] M. Bauer, D. Collin, U. Iemma, K. Janssens, F. Marki and U. Müller, "COSMA - A European approach on aircraft noise annoyance research," in *Internoise 2014*, Melbourne, Australia, 2014.

[40] A. Sahai and D. Simons, "Quantifying the audible differences in measured and auralized aircraft sounds," in *22nd AIAA/CEAS Aeroacoustics conference*, Lyon, France, 2016.

[41] A. Minard, "Signal-based indicators for predicting the effect of audible tones in the aircraft sound at takeoff," in *Inter-Noise 2016*, Hamburg, Germany, 2016.

[42] Ecma International, "Standard ECMA-74, 13th edition: Measurement of airborne noise emitted by information technology and telecommunications equipment," Ecma International, Geneva, 2015.

[43] S. Rizzi, C. Burley and R. Thomas, "Auralization of NASA N+2 Aircraft Concepts from System Noise Predictions," Lyon, 2016.