Quantifying the audible differences in measured and auralized aircraft sounds

Sahai, Abhishek; Simons, Dick

DOI
10.2514/6.2016-2909

Publication date
2016

Document Version
Accepted author manuscript

Published in
22nd AIAA/CEAS Aeroacoustics Conference

Citation (APA)
Sahai, A., & Simons, D. (2016). Quantifying the audible differences in measured and auralized aircraft sounds. In *22nd AIAA/CEAS Aeroacoustics Conference: Lyon, France* Article AIAA 2016-2909 American Institute of Aeronautics and Astronautics Inc. (AIAA). https://doi.org/10.2514/6.2016-2909

Quantifying the audible differences in measured and auralized aircraft sounds

Abhishek K. Sahai[1] and Dick G. Simons[2]
Aircraft Noise and Climate Effects (ANCE) Research Group, Faculty of Aerospace Engineering,
Delft University of Technology, Delft, 2625HS, The Netherlands

This paper aims to present a way with which audible differences in measured and auralized (i.e. aurally simulated) aircraft noise can be quantified. The purpose of the study is to find a means with which the subjective differences between measured and synthesized sounds can be expressed in an objective manner. The quantification would firstly enable developers of auralization technology to identify more concretely in which aspects the differences exist, in order to make the auralizations sound more realistic. The quantification would secondly aid in developing a means of distinguishing between aircraft sounds in general, beyond the conventional metrics of A-weighted level (dBA) or Effective Perceived Noise Level (EPNL). Such a capability can allow target functions to be developed with which aircraft can be optimized for specific, more acceptable sounds. As used widely in other industries such as the automotive sector, use of sound quality metrics is made to quantify the differences in the quality of the sounds. The comparison is carried out in terms of both conventional and sound quality metrics for the audio of a reference aircraft, which has been measured and auralized over the same flight paths at a noise monitoring station in the airport vicinity.

Nomenclature

A = Amplitude
ϕ = Phase
f = Frequency
f_s = Sampling frequency
N_{block} = Number of samples in time-block
p_0 = Pressure at threshold of hearing
N_{noy} = Overall perceived noisiness
n = Perceived noisiness in a 1/3 octave frequency band
C = Tone correction
D = Duration correction
z = Critical band rate
N = Loudness
N' = Specific Loudness
L = Excitation level
S = Sharpness

I. Introduction

THE noise produced by aircraft has traditionally been expressed in a specific overall value such as A-weighted level (dBA), Sound Exposure Level (SEL) and Effective Perceived Noise Level (EPNL). Any differences in the

---

[1] Assistant Professor, ANCE Research Group, Faculty of Aerospace Engineering, Delft University of Technology, 2625HS Delft, The Netherlands.

[2] Full Professor, ANCE Research Group, Faculty of Aerospace Engineering, Delft University of Technology, 2625HS Delft, The Netherlands.

noise two aircraft produce have therefore also till now been expressed in these commonly used metrics by the aerospace community. Problems arise however when two aircraft have very similar weighted levels and EPNL values yet their sounds are noticeably different when heard by observers during flyovers. It is known that traditional metrics such as dBA and others that use it as a basis, do not suffice in clearly distinguishing between aircraft sound signatures and also do not correspond well to the actual perceived annoyance due to aircraft noise [1]. The same can also be said to some extent for the EPNL metric [2], particularly in capturing differences in tonal content of aircraft noise [3], [4]. These conventional metrics can therefore prove to be lacking in capturing important differences in aircraft sounds, such as the prominence of tonal noise in relation to broadband noise, fluctuations in loudness over time (particularly important for propeller based aircraft) or the ratio of high to low frequency noise for instance. The ability to clearly distinguish between aircraft sounds is of fundamental importance to reducing the adverse effects due to aircraft noise experienced by residents, and for any efforts aimed at increasing the acceptance of aircraft noise. Any noise reductions achieved solely by focusing on metrics that do not fully capture the individual aspects of aircraft sounds may not yield a reduction in the actual annoyance experienced by residents or an increase in the acceptance of aircraft noise.

The need to be able to express differences in aircraft sounds in an improved and objective way was particularly highlighted when listening to a comparison of measured and synthesized aircraft noise audio of the same aircraft [5]. It was observed in this work by Arntzen et al. that audible differences clearly existed even though the noise impact values in dBA_{max} were 0-3 dBA and in SEL 0-4 dBA. Some of the audible differences noted were for instance that the synthesized tones were too prominent, the low frequency noise was underpredicted and that there was more turbulence in the recordings due to the presence of wind-gusts. These differences were not captured by the A-weighted metrics. Also other studies on the auralized sounds of future Counter Rotating Open Rotor (CROR) engines [6] showed that noise reductions due to improved blade design could be expressed in EPNL values but these could not clearly specify audible changes in the quality of the sounds. It becomes clear that when differences in audio cannot be expressed objectively in quantifiable numbers, they have to be expressed subjectively, which does not provide a practical capability for optimization or improvement of the sounds. It is therefore reasoned that if these differences could be objectively quantified, then the quality of the auralized sounds could be improved. It could then be investigated more concretely which audio components have to be adjusted or modified in order to bring the synthetic sound closer to the measured sound. Since the sound produced by the aircraft is directly linked to different noise sources on the aircraft (such as fan tones, broadband jet or airframe noise), it could also be investigated which source noise component would require a higher modeling accuracy in order to remove the respective difference in sound.

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 characterization of the aircraft sound signatures would then serve the second more overall goal of aircraft design for optimal sound, whereby aircraft designs could be optimized to meet required, more acceptable target sound signatures. This would shift the focus from ‘low-noise’ aircraft design towards ‘low-annoyance’ aircraft design, an approach which focuses on the optimization of aircraft sounds reaching the observers rather than of overall levels [3], [4], [7].

The paper has been divided into five main sections. The aircraft noise measurement and auralization methodology has been explained in more detail in [5] and is briefly recapitulated in Section II; the sound assessment methodology in both conventional and sound quality metrics is explained in Section III; the comparison of the aircraft noise assessment in conventional and sound quality metrics is performed in Section IV and Section V presents the conclusions of the current work.

II. Noise measurement and auralization methodology

For comparison of the measured and auralized aircraft noise audio, a reference aircraft is chosen, in this case the Boeing 747-400 equipped with four CF6-80C2 engines. The comparison is made for four takeoff flight paths of the 747-400 at a noise monitoring point near Schiphol airport Amsterdam, located 3.8km in front and 400m to the right of the runway in the takeoff direction. The noise has been both measured and auralized at this location. The average weather conditions for the day of the measurement were a temperature of 10°C, pressure of 999.6hPa and relative humidity of 96%. The noise monitoring station was located on a grassy field and to minimize the effect of ground reflections, the microphone was located at a height of 10m. As will be seen in the measured spectrograms in Section

IV, relatively strong wind-gusts were present during the day of the measurements, which are observed as vertical spikes in the spectrograms. Flight paths 1 to 3 were quite similar in the trajectory followed, with the aircraft flying closer to the monitoring station than during flight path 4. For further details regarding the flight paths and measurements thereof, the reader is referred to [5].

In order to create auralized aircraft noise of the measured flight paths, the inputs for the fan, jet and airframe source noise models had to be simulated over the flight paths. 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 GSP model [8], based on the required thrust for the aircraft takeoff weight, lift and drag (using a relevant lift-drag polar for the aircraft). The source noise models used are based on NASA's Aircraft Noise Prediction Program (ANOPP) [9], which include the model of Heidmann [10] for fan and compressor noise, Stone [11] for jet noise and Fink [12] for airframe noise. Since combustor and turbine noise are not dominant during takeoff, their simulations were left out of the prediction and analysis. These source noise models are semi-empirical in nature and although they do not allow a hundred percent match to measured data, particularly for more modern aircraft of today, they do provide a generic noise prediction capability for all conventional aircraft and engines, besides being not very computationally expensive.

The simulated fan tones and broadband jet as well as airframe noise are then synthesized at the source. For auralizing tonal noise, an *additive synthesis technique* [13], [14], [15] has been used, which is shown via Eqs. 1 and 2.

$$s_i(t) = A_i \cos(\phi_i(t) + \phi_0) \tag{1}$$

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

Here each tone is constructed as a cosine wave having amplitude A_i, instantaneous phase ϕ_i and an initial phase ϕ_0, set here as a random phase offset to produce a more realistic and less coherent sound. The instantaneous phase ϕ_i is calculated from the instantaneous frequency f_i. By constructing the individual tones using this technique, the total tonal component can be constructed via a simple summation of the individual tones.

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 occur at multiples of the low pressure shaft speed (*N1*). Although the interaction tones have their individual magnitudes specified by the method of Heidmann, the magnitudes of the buzz-saw tones are provided over a 1/3 octave spectral division. By knowing in which 1/3 octave band the 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, a technique making use of white noise is used. The 1/3 octave source noise spectra provided by the models of Stone for jet noise and Fink for airframe noise are firstly converted to a narrowband noise spectrum. White noise is generated in the frequency domain and is 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 noise reaching the observer changes due to the continuously changing emission angle between the aircraft and the observer and also due to changes in the aircraft's thrust setting and/or high-lift device setting. This implies that the convolution of the white noise and narrowband spectrum changes with time and can result in audible artifacts as the aircraft flies by. To avoid these artifacts, an Overlap Add (OLA) technique is employed to combine the signals with an overlap after windowing them using a Hanning window.

As against auralizing the propagated aircraft noise on the ground as done in [15], [16], the noise has been synthesized for the audio comparison firstly at the source. In order to reproduce the noise impact at the noise monitoring point, the propagation and flight effects are subsequently applied to the source noise as various gains and filters. As the source noise models are intended to predict far-field noise, the whole aircraft along with the engines is regarded as a point source for applying the propagation effects of spherical spreading, atmospheric absorption according to ISO-9613-1:1993 [17] and ground reflection according to Chien-Soroka theory [18] using Delany and Bazley's [19] ground impedance model. Since the microphone is however located at a height of 10m, the effects of ground reflection are in this case minimal. The Doppler shift is further applied to account for the moving source effect 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 [16]. In this regard, the VDL makes use of a spline interpolation in order to avoid aliasing effects, if the time-delay results in a non-integer retarded emission time.

III. Noise assessment methodology in conventional and sound quality metrics

As mentioned in Section I, the measured and auralized aircraft sounds will be assessed and compared in this paper in conventional metrics as well as in the sound quality metrics. Derived from the field of psychoacoustics, five metrics which focus on individual spectral and temporal characteristics of sounds are recognized as sound quality metrics – loudness, tonality, roughness, sharpness and fluctuation strength. The loudness metric is further divided into stationary and time-varying loudness. The dBA_{max}, SEL and EPNL metrics, as well as the stationary loudness and sharpness metrics have been implemented by the authors as part of an aircraft noise audio assessment module. The ultimate goal of developing this module is to have a capability which could be automated and integrated in an aircraft design – auralization – sound assessment chain, providing the aircraft design for optimal sound or low-annoyance aircraft design possibility in the future. The remaining sound quality metrics have been applied to the analysis of the audio files using Bruel and Kjaer's Pulse Reflex software. Since the assessment in terms of the sound quality metrics will be discussed in detail in the next section, it is helpful to briefly describe what these metrics measure and which method their calculation has been based on.

- *Loudness* is the subjective perception of the overall magnitude of a sound [22]. For complex sounds, effects such as the masking of high frequency noise components from lower frequency noise are important in determining the overall perceived loudness. Stationary loudness can be applied to non-impulsive sounds. For impulsive sounds, time-varying loudness measures loudness changes for very small durations of 1ms and additionally incorporates temporal masking effects [23]. The method of Zwicker has been used in this paper to determine the stationary and time-varying loudness.
- *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 isn't standardized and several software such as Pulse Reflex use Terhardt's tonality metric, which estimates the virtual pitch of a sound containing multiple tonal components [25]. 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 does not determine the pitch salience but considers the influence of frequency, bandwidth and SPL excess of tonal and narrowband components, as well as their loudness relative to the overall loudness to determine the tonality of a sound. Aures' tonality metric has been applied by the authors in previous studies on low-annoyance aircraft design [3], [4], [7] and will also be added to the audio assessment module in the future.
- *Sharpness* is a measure of the high frequency content of a sound and can indicate to what extent a sound is dominated by high frequency components [26]. This attribute can potentially aid in distinguishing between aircraft sounds in this paper, by focusing on the ratio of high to low frequency aircraft noise. The sharpness metric has been computed using the method of von Bismarck for this paper.
- *Roughness* is the sensation produced by sounds containing fast loudness fluctuations in the order of 50-90 cycles per second [22]. Roughness is of less relevance for the sounds of turbofan engines but can be an important parameter for propeller based aircraft. One aspect of turbofan engine noise that is however comparable to the noise of propeller based aircraft are the buzz-saw tones from the fan. The influence of these buzz-saw tones may be captured by the roughness metric in this paper, which has been applied using Zwicker's method in Pulse Reflex.
- *Fluctuation strength* is a measure of slow fluctuations in loudness of the order of 1-16 cycles per second. This metric may help in distinguishing between the various aircraft sounds if any slow fluctuations in the pressure variation over time are present. The Pulse Reflex software applies the fluctuation strength metric using Zwicker and Fastl's method [22].

The methodology with which the conventional metrics as well as the sound quality metrics of stationary loudness and sharpness have been implemented in the audio assessment module can now be explained in this section.

The samples of the audio file to be assessed are firstly divided into blocks, in time-steps of dt_{block} = 0.125 seconds. For each block with number of samples N_{block}, the Power Spectral Density (PSD) is computed using the Fast Fourier Transform (FFT) for each block according to:

$$PSD_{block}(f) = 2\left(\frac{dt^2}{dt_{block}}\right)|FFT(f)|^2 \tag{3}$$

Here f is defined in steps of $\Delta f = f_s/N_{block}$ from 0 to $f_s/2$, i.e. till the Nyquist frequency, with f_s being the sampling frequency of 44.1 kHz and dt is the sampling interval equal to $1/f_s$. The term dt^2 has to be multiplied with the square of the FFT as computed by Matlab in order to get the correct values of the PSD using Eq. 3. From the PSD values computed at each frequency f in each block, the Sound Pressure Level (SPL) at each frequency in each block can be computed as:

$$SPL_{block}(f) = 10log_{10}\left(\frac{PSD_{block}(f)\Delta f}{p_0^2}\right) \tag{4}$$

In Eq. 4, p_0 is the pressure at the threshold of hearing, $2\cdot10^{-5}$ Pa. As several metrics including EPNL and stationary loudness require 1/3 octave spectra as inputs, a 1/3 octave filter is then applied to get the levels at 1/3 octave band center frequencies.

A. Methodology for assessment in conventional metrics

In order to calculate the maximum A-weighted level, dBA_{max}, A-weighting is firstly applied to the levels at each 1/3 octave frequency. In a next step, the Overall A-weighted SPL (OASPL) is calculated for each block by logarithmically adding the SPL values at each frequency, thereby yielding an overall A-weighted level vs time history of the aircraft flyover. The maximum value from this history is therefore the dBA_{max} value. In order to calculate the SEL, the 10 dBA down value from dBA_{max} is determined and the SEL value is computed according to Eq. 5, where t_1 and t_2 specify the time-interval for which the noise impact is 10 dBA below dBA_{max}:

$$SEL = 10log_{10}\left[\int_{t_1}^{t_2} 10^{\frac{dBA(t)}{10}} dt\right] \tag{5}$$

The EPNL metric is calculated according to the algorithm outlined by the Federal Aviation Administration (FAA) of the United States [20] and requires firstly the calculation of the Perceived Noise Level (PNL). The PNL is an annoyance based metric, developed by the FAA in the 1960s to certify aircraft for noise and makes use of equal noisiness curves. The ‘noisiness’ is expressed linearly in the unit of noy according to Eq. 6 and logarithmically in the unit of Perceived Noise decibel, PNdB according to Eq. 7:

$$N_{noy} = n_{max} + 0.15\left[\sum_{i=1}^{24}(n_i) - n_{max}\right] \tag{6}$$

$$PNL = 40 + \frac{10\, log_{10} N_{noy}}{log_{10} 2} \tag{7}$$

Here, n_{max} is the maximum noy value in the whole 1/3 octave spectrum and n_i is the noy value at the i^{th} 1/3 octave frequency.

The fact that the presence of discrete tones in otherwise broadband sounds makes the sound perceptually more annoying [21], led to the addition of a tone correction to the calculated PNL value. The Tone-corrected Perceived Noise Level (PNLT) metric therefore adds a tonal penalty to the PNL values if the presence of a strong protruding tone in the spectrum is detected. This is done according to Eq. 8, where C is the tone-correction factor:

$$PNLT = PNL + C \tag{8}$$

Incorporation of a duration correction is the final addition to the PNL metric that ultimately yields a single value measure for the annoyance caused by aircraft noise for a single event. Similar to the time integration performed for the SEL metric, PNLT values 10 PNTdB below the maximum PNLT value are integrated for the aircraft movement and then normalized for a 10 second interval to calculate the duration correction D, according to Eq. 9. This leads to a final EPNL value using Eq. 10.

$$D = 10log_{10}\left[\sum_{k=0}^{2d} 10^{PNLT(k)/10}\right] - PNLTM - 13 \tag{9}$$

$$EPNL = PNLTM + D \tag{10}$$

In Eq. 9, k is the index for each time-step, given a standard value of 0.5 seconds for the certification process, d is the time-interval for which the tone-corrected perceived noise level value is 10 PNTdB below the maximum value *PNLTM* for the movement in consideration. The value 13 is obtained due to normalization for the 10 second interval i.e. $10\log_{10}(0.5/10)$. Since the time-step for each block dt_{block} has been taken to be 0.125s for the current work, the time-integration for the EPNL calculation is performed for each fourth block, in order to get the correct EPNdB values.

B. Methodology for assessment in sound quality metrics

The stationary loudness metric has been implemented in the audio assessment module using Zwicker's loudness calculation method as described in ISO532-B and DIN 45631 [27]. This method is based on 1/3 octave spectra provided as inputs and involves three primary calculation steps to determine the loudness perception of a sound – modeling the transmission of the acoustic signal through the outer and middle ear, estimation of the spectral masking effects to determine an unmasked specific loudness pattern, and an integration of the specific loudness pattern to get an overall loudness value in sone. In order to model the transmission characteristics of the human ear and the spectral masking effects, Zwicker's method divides the audible frequency range into a series of frequency bands called *critical bands*, which are given the unit of Bark. A frequency f in Hz can be converted to the Bark scale using the critical band rate z, which has a value from 0 to 24 Bark:

$$z = 13\tan^{-1}\left(0.76\frac{f}{1000}\right) + 3.5\tan^{-1}\left(\frac{f}{7500}\right)^2 \tag{11}$$

The *critical bandwidth* of each critical band corresponds to the frequency resolution of the human ear. The corresponding critical bandwidth, *CBW* can then be calculated as:

$$CBW = 25 + 75\left(1 + 1.4\left(\frac{f}{1000}\right)^2\right)^{0.69} \tag{12}$$

The sound, expressed in 1/3 octave spectra, produces an excitation over several critical bands. Using the excitation level $L_E(z)$ in each critical band and by checking if the excitation level lies above the level at the threshold of hearing $L_{TQ}(z)$ in each critical band, the specific loudness N' in sone/Bark, is then calculated as:

$$N'(z) = 0.0635 \cdot 10^{0.025L_{TQ}(z)}\left[\left(0.75 + 0.25 \cdot 10^{0.1(L_E(z) - L_{TQ}(z))}\right)^{0.25} - 1\right] \tag{13}$$

The specific loudness calculated using Eq. 13 has so far not considered any masking effects and is referred to as the main loudness in each critical band. The excitation caused by any sound focused on one critical band however spreads over multiple critical bands and has a specific masking pattern, whereby the spread of excitation is larger towards higher frequencies than towards lower frequencies. The spectral masking patterns as a function of the critical bands were determined by Zwicker by playing pure tones surrounded by narrowband noise at several frequencies to test audiences. By checking in each critical band if the excitation is masked or not by the excitation from a sound component focused on another critical band, an unmasked main loudness value in each critical band is determined. The unmasked main loudness is then used to calculate the total loudness value N in sone, which corresponds to calculating the area under the unmasked specific loudness pattern as:

$$N = \int_0^{24} N'(z)dz \tag{14}$$

The sharpness metric has been implemented in the audio assessment module based on the method of von Bismarck, as mentioned earlier. The sharpness calculation method makes use of a weighting function $g(z)$, which was determined using psychoacoustic tests carried out by von Bismarck and weighs all spectral content at or above 16 Bark (i.e. 2700 Hz and above) more heavily:

$$g(z) = \begin{Bmatrix} 1 & z \leq 16 \\ 0.066e^{0.171z} & z > 16 \end{Bmatrix} \tag{15}$$

For a given 1/3 octave band spectrum, the overall sharpness value S in the unit of acum is then calculated using Eq. 16, where the constant c has a value of 0.11.

$$S = c\left[\frac{\int_0^{24} g(z)N'(z)z\,dz}{N}\right] \tag{16}$$

IV. Comparison of measured and auralized aircraft noise

A. Measured and auralized aircraft noise audio spectrograms

Figures 1 and 2 show the spectrograms of measured and auralized aircraft noise for flight path 1, as shown by Arntzen et al. in [5]. The vertical lines or 'spikes' mentioned earlier due to turbulent wind-gusts can be seen in the measured spectrogram shown in Fig. 1. It is evident by looking at the spectrograms that the synthesized audio in Fig. 2 is much cleaner and lacks the significant turbulence that can be noticed in the measured audio's spectrogram. Also the fundamental fan interaction tone, beginning at around 2800 Hz, is seen to be much cleaner and more pronounced in the synthesized audio's spectrogram in Fig. 2. The prominence of the fundamental fan tone is in this case not only due to the fan tonal noise intensity being overpredicted in the source noise models, but also due to the fact that less broadband noise surrounds the fan tones. The buzz-saw tones are also overpedicted in the synthesis and can be seen quite clearly as the numerous horizontal lines in the first half of the spectrogram in Fig. 2. 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 spectrogram in Fig. 2. 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 Fig. 1. This is done in order to minimize data-storage owing to the high volume of air-traffic. The audio information has therefore been provided also for the auralized noise only till a frequency of 3500 Hz.

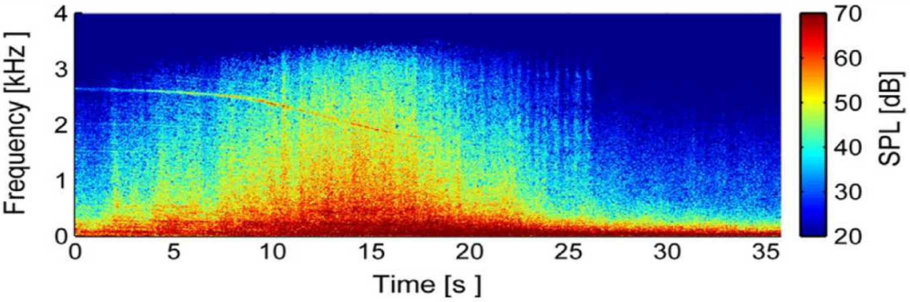

Figure 1: Measured takeoff spectrogram for flight path 1 from [5]

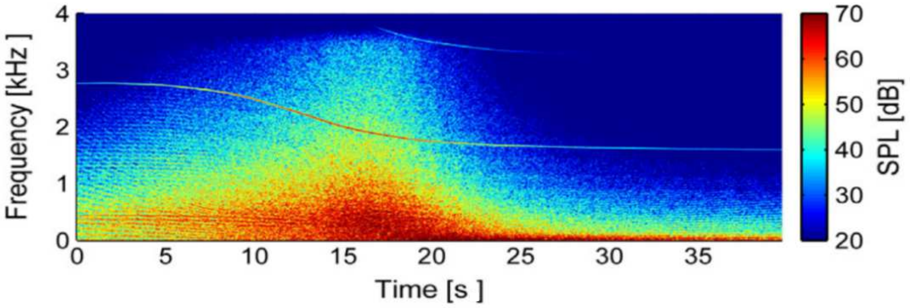

Figure 2: Auralized takeoff spectrogram for flight path 1 from [5]

Fig. 3 and Fig. 4 show the measured and auralized spectrograms for flight path 4, for which the aircraft flies at a larger distance from the noise monitoring station. The aircraft in flight path 4 therefore has a flatter SPL vs. time variation than what is observed in flight path 1 and in the other flight paths shown in [5]. A very noticeable difference for flight path 4 is in the vertical spikes in the spectrograms, which appear to be spread out over time compared to flight path 1, for which they lied closer together. The fundamental fan tone also appears to be less pronounced for the synthesized noise in flight path 4 than what was seen for flight path 1, except at the beginning and end of the flyover. The remaining differences observed for flight path 1, such as the low frequency noise having a higher intensity in the measurements and the buzz-saw tones being overpredicted in the synthesis, are also observed in the spectrograms presented for flight path 4. The next sub-section attempts to quantify these differences in the individual aircraft noise characteristics, using both the conventional and sound quality metrics.

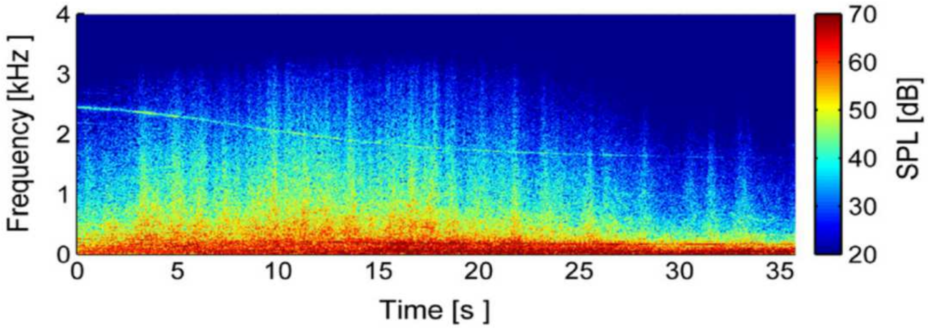

Figure 3: Measured takeoff spectrogram for flight path 4 from [5]

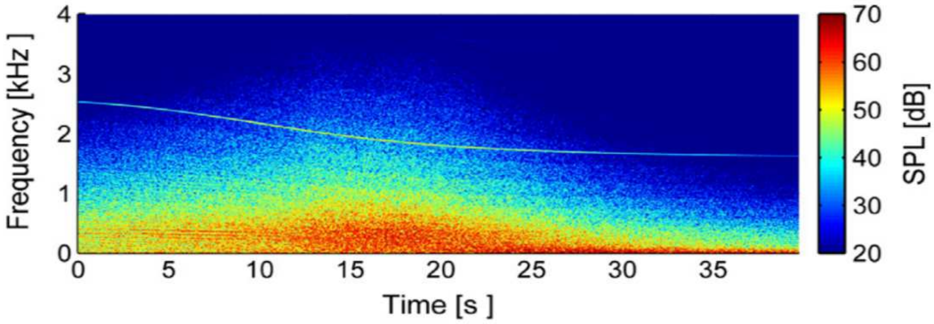

Figure 4: Auralized takeoff spectrogram for flight path 4 from [5]

B. Measured and auralized aircraft noise audio assessment

1. Assessment in conventional metrics

The differences in terms of the conventional metrics are presented in Table 1 in absolute values of dBA or EPNdB, as these metrics are logarithmic in nature. The absolute dBA and SEL metric values are lower than what was observed in [5] as the original wav files had to be reproduced for the current work from mp3 format, which reduces the maximum intensities of audio files. The differences in the metric values between the measured and auralized sounds are nonetheless of the same order as to what was observed in [5], thereby indicating that the relative differences presented in this paper remain valid. By comparing the measured and synthesized sounds objectively in conventional metrics in Table 1, it can be seen that in terms of the dBA values, the differences in the dBA_{max} values lie within a range of 0.6-3 dBA between the measured and synthesized sounds and those in the SEL values within 0.5-3.8 dBA. These differences are on the whole small and also do not indicate in which way the aircraft sounds differ. A difference of 3dBA is therefore an indicator that a reasonable difference in the sounds exists but it does not clarify in which way the sounds are different and what the probable cause of this difference could be. The conventional metric analysis was extended to also include the EPNL metric according to the methodology explained in Section III. The differences in EPNdB are on the whole larger than what is observed for the dBA based metrics for the first three flight paths. The EPNL metric values differ for flight path 1 by 1.3 EPNdB, which is twice the difference indicated by dBA based metrics. The larger difference may be due to the more prominent fan tone in the synthesis, which is captured to some extent by the tonal penalty of the EPNL metric. For flight paths 2 and 3, the differences are of similar order to the dBA_{max} and SEL metrics, with a 3.4-3.8 EPNdB difference indicating a significantly different sound. The difference for the fourth flight path is however very small, in contrast to the dBA_{max} and SEL metrics. As was mentioned for the dBA based metrics, the EPNL metric can indicate overall differences in aircraft sounds but it does not provide clear information regarding the differences in the individual aircraft sound characteristics.

Table 1: Comparison of measured and synthesized aircraft sounds in the conventional metrics of dBA_{max}, SEL and EPNL

Flight Path	Measured dBA_{max} [dBA]	Synthesized dBA_{max} [dBA]	Δ [dBA]	Measured SEL [dBA]	Synthesized SEL [dBA]	Δ [dBA]	Measured EPNL [EPNdB]	Synthesized EPNL [EPNdB]	Δ [EPNdB]
1	69.2	68.6	-0.6	74.9	74.4	-0.5	80.1	78.8	-1.3
2	71.0	68.5	-2.5	76.0	73.4	-2.6	81.7	78.3	-3.4
3	70.3	67.3	-3.0	77.7	73.9	-3.8	80.9	77.1	-3.8
4	66.3	68.8	+2.5	75.6	76.5	+0.9	81.1	80.8	-0.3

2. Assessment in sound quality metrics

Tables 2 and 3 show the comparison of measured and synthesized audio in terms of the sound quality metrics. N_5 in Table 2 refers to the time-varying loudness exceeded for 5% of the time of the flyover duration and K refers to the tonality calculated using Terhardt's method; S in Table 3 refers to the sharpness, R to the roughness, and FS to the fluctuation strength. As the sound quality metrics presented in Tables 2 and 3 are linear in nature and their use for aircraft noise analysis is relatively new, the differences are shown as relative values rather than absolute values in order to present the differences in sound quality in a more relatable and understandable form.

Table 2: Comparison of measured and synthesized aircraft sounds in stationary loudness, time-varying loudness and tonality metrics

Flight Path	Measured stationary N [sone]	Synthesized stationary N [sone]	Δ [%]	Measured time-varying N_5 [sone]	Synthesized time-varying N_5 [sone]	Δ [%]	Measured K [-]	Synthesized K [-]	Δ [%]
1	15.62	14.07	-9.9	21.7	20.2	-6.9	0.0496	0.0616	+24.2
2	17.1	13.35	-21.9	23.67	20.36	-14.0	0.0418	0.0711	+70.1
3	18.11	13.13	-27.5	24.66	19.22	-22.3	0.0377	0.0632	+67.6
4	16.15	15.01	-7.1	20.31	19.23	-5.3	0.0519	0.0521	+0.4

The analysis of the audio assessment in sound quality metrics can be performed firstly for the stationary loudness, time-varying loudness and tonality metrics, presented in Table 2. It can be observed that the differences between the measured and auralized audio files are much greater in terms of the loudness metrics when expressed linearly in sone. The differences of 0.5-0.6 dBA for flight path 1 in the dBA_{max} and SEL metrics are seen as differences up to 10% in stationary loudness. The relative differences for flight paths 2 and 3 are much higher, at 21.9 and 27.5% respectively, indicating that the measurements are on the whole louder than the synthesis, which contributes to the larger differences in the dBA based and EPNL metrics. The differences in loudness when expressed logarithmically as the loudness level in phon however are -1.51 phon, -3.58 phon, -4.64 phon and -1.06 phon for each of the four flight paths respectively, which are closer to the EPNdB differences than to the dBA differences. The comparison of the loudness and dBA differences indicates that although the dBA metric is based on the 40 phon equal loudness level curve, the loudness metric still provides clearer information regarding the differences in overall intensity of aircraft sounds. It can also be observed that the loudness differences are being captured by the EPNL metric. The knowledge from comparing the loudness values can be beneficial in improving the quality of the auralized aircraft noise, that increasing the overall loudness by the amounts specified by the loudness metric could make the synthetic sounds be perceived as being more similar to the measured sounds.

Table 2 shows that the largest differences are seen to occur for the tonality metric. It was apparent by looking at the spectrograms in Figs. 1-4 that the tonal content in the auralized audio had been overpredicted and was more pronounced. These subjective differences are seen as very large relative differences of up to 70% for the perceived tonality of the synthesized fan tones, when compared to the measured fan tones. Part of the reason for the much higher perceived tonality is that the buzz-saw tones are much more prominent in the synthesized noise than in the measured noise. To a large extent however, the differences are due to the fundamental fan interaction tone being overpredicted and more pronounced in relation to the surrounding broadband noise. Quantification of the subjectively perceived differences in the tonal content via the tonality metric can help the developers of auralization technology 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 adjustment of the tonal content 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 modified to minimize the significant differences in perceived tonality of the simulated aircraft noise. Comparing the differences in tonality to the EPNdB differences, it can be observe that the additional differences in the sounds due to the tonal content are not clearly reflected by the EPNL metric. This indicates as well that the tonal content via the tonal penalty is not being sufficiently captured by the EPNL metric, which is seen to have lower values for all flight paths similar to the loudness metrics but does not show any increase in values due to the overpredicted tonal content in the synthesized sounds.

Table 3: Comparison of measured and synthesized aircraft sounds in sharpness, roughness and fluctuation strength metrics

Flight Path	Measured *S* [acum]	Synthesized *S* [acum]	Δ [%]	Measured *R* [asper]	Synthesized *R* [asper]	Δ [%]	Measured *FS* [vacil]	Synthesized *FS* [vacil]	Δ [%]
1	0.829	0.859	+3.6	1.46	1.67	+14.4	1.21	1.22	+0.8
2	0.884	0.879	-0.6	1.49	1.59	+6.7	1.37	1.41	+2.9
3	0.878	0.872	-0.6	1.46	1.51	+3.4	1.27	1.31	+3.1
4	0.748	0.784	+4.8	1.29	1.2	+7.0	1.02	0.819	-19.7

Looking at Table 3, it can be seen that the differences in the sharpness metric values are not very high between the measured and auralized aircraft noise, for all of the flight paths. It was evident from the spectrograms presented in Figs. 1-4 that the measured aircraft noise clearly has more low frequency noise than the synthesized aircraft noise. This effect has therefore not been captured by the sharpness metric, in its current form, with higher weightage given to high frequency noise lying above 2700 Hz. One possible reason for this could be the low-pass filtering carried out at the noise monitoring station whereby noise above 3500 Hz had to be filtered out. It is possible that some of the higher frequency differences were therefore filtered out as well. It is known however that aircraft noise above 3500 Hz when propagated on the ground is in general of low intensity, due to absorption of the higher frequency components by the atmosphere. Furthermore, the sharpness metric was also found to have the lowest correlation with the perceived annoyance due to aircraft noise by both [1] and [2], implying that changes in high frequency noise above 2700 Hz did not cause any noticeable change in the perception of aircraft noise. Based on these two observations, it can be said that the sharpness metric would need to be modified in order to be more suitable for application to aircraft noise. One possibility may be to apply the higher weightage from a lower frequency than 2700 Hz, where there is more aircraft noise present such as close to 1000 Hz or lower. It could then be investigated if the identified differences in low frequency content can be captured with the reduced high frequency threshold.

The roughness and fluctuation strength metrics focus on the temporal variations of the sounds, rather than the spectral characteristics. It was seen in the spectrograms that buzz-saw noise had been overpredicted in the synthesis. Buzz-saw noise involves several closely spaced tones, which result in fast amplitude fluctuations in the aircraft sound during takeoff. These fast fluctuations in pressure amplitude result in fast fluctuations in loudness, which is reflected in the higher roughness values for the synthesized aircraft noise. The relative differences in roughness values are highest for flight path 1 at 14% and are also significant for flight path 2 and 4. This would indicate that the buzz-saw tones' intensity has to be reduced at the source, similar to the reduction of the fan interaction tones, to improve the roughness perception of the synthesized sounds and make them approach the measured aircraft sounds.

The fluctuation strength metric, as mentioned earlier, aims to capture slower fluctuations in loudness over time. Looking at Table 3, it can be seen that the difference in slow loudness fluctuations is not captured for the first three flight paths, with the synthesized values lying within 3% of the measured values. For flight path 4 however, the metric values show that the synthesized aircraft noise has 20% less slow loudness fluctuations over time compared to the measured aircraft noise. Looking back at Figs. 3 and 4, it can be seen that this likely refers to the vertical spikes occurring due to turbulent wind-gusts in the measurements, which could not be modeled in the synthesis. A suggestion for the improvement of the synthesis for flight path 4 would therefore be to introduce these frequency independent wind-gusts periodically into the synthesis, to bring the synthesized fluctuation strength closer to that of the measurement. More analysis is however required as to why the temporal fluctuations due to the turbulent wind-gusts were not captured by the fluctuation strength metric for all of the flight paths.

Conclusions

It is observed that the differences between measured and synthesized aircraft noise audio are either small when expressed in the conventional dBA_{max}, SEL and EPNL metrics or do not clearly indicate which aircraft sound characteristics are causing the differences in their values. In contrast, these differences are amplified when expressed in the sound quality metrics and can be expressed in a clearer way by focusing on different aircraft noise characteristics. It was found that the synthesized aircraft noise was overall quieter than the measured noise, with differences in sone of up to 27.5% observed in stationary loudness and 22.3% in time-varying loudness. The differences expressed logarithmically in phon are on the whole larger than the dBA_{max} values and are of similar order to the EPNL differences. The most significant differences were observed in the tonality of the sounds, with the auralized sounds being perceived as up to 70% more tonal than the measured sounds. A reduction of the fan tonal intensity and its prominence would aid in bringing the synthesis closer to the measurements. Another aspect that would improve the comparison of the synthesized aircraft noise with the measurements would be the inclusion of temporal effects such as the turbulence due to wind-gusts. These were captured by the fluctuation strength metric for flight path 4, which showed that the auralized noise had 20% less slow loudness fluctuations than the measured noise. The wind-gusts however could not be distinguished for the other flight paths by the fluctuation strength metric and more analysis would be needed to find the right metric to capture the wind-gusts for all cases. Further analysis would also be beneficial with regards to the applicability of the sharpness metric to aircraft noise analysis, which is currently unable to capture differences in aircraft noise frequency content that are visible in the spectrograms.

The analysis presented in this paper highlights the need to look beyond the conventional metrics used today in order to be able to distinguish between aircraft sounds. The ability to objectively quantify differences between aircraft sounds and their individual characteristics can allow improvement of the aircraft noise auralization capability and also aid in increasing the accuracy of the source noise models being used in the synthesis. Distinguishing clearly between aircraft sounds is essential in order to improve the quality of aircraft sounds reaching the residents and in this way increase the acceptance of aircraft noise. From the analysis carried out in this paper, the use of sound quality metrics provides a clearer way of doing this than the conventional metrics used today.

References

[1] 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.

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

[3] 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.

[4] A. K. Sahai, M. Snellen, D. G. Simons and E. Stumpf, "Lowering the perceived annoyance due to aircraft noise through aircraft design optimization," *AIAA Journal of Aircraft,* no. (Submitted for publication), 2016.

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

[6] 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.

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

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

[9] W. E. Zorumski, "Aircraft noise prediction program theoretical manual," NASA TM 83199, 1982.

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

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

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

[13] M. Allen, "Analysis and synthesis of aircraft engine fan noise for use in psychoacoustic studies," Virginia Polytechnic Institute and State University, Master thesis, 2012.

[14] 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.

[15] 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.

[16] S. Rizzi and B. Sullivan, "Synthesis of virtual environments for aircraft community noise impact studies," in *11th AIAA/CEAS Aeroacoustics Conference*, Monterey, CA, USA, 2005.

[17] 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.

[18] 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.

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

[20] 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.

[21] 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.

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

[23] 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.

[24] W. Aures, "Procedure for calculating the sensory euphony of arbitrary sound signals," *Acustica,* vol. 59, pp. 130-141, 1985.

[25] 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.

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

[27] 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. vol