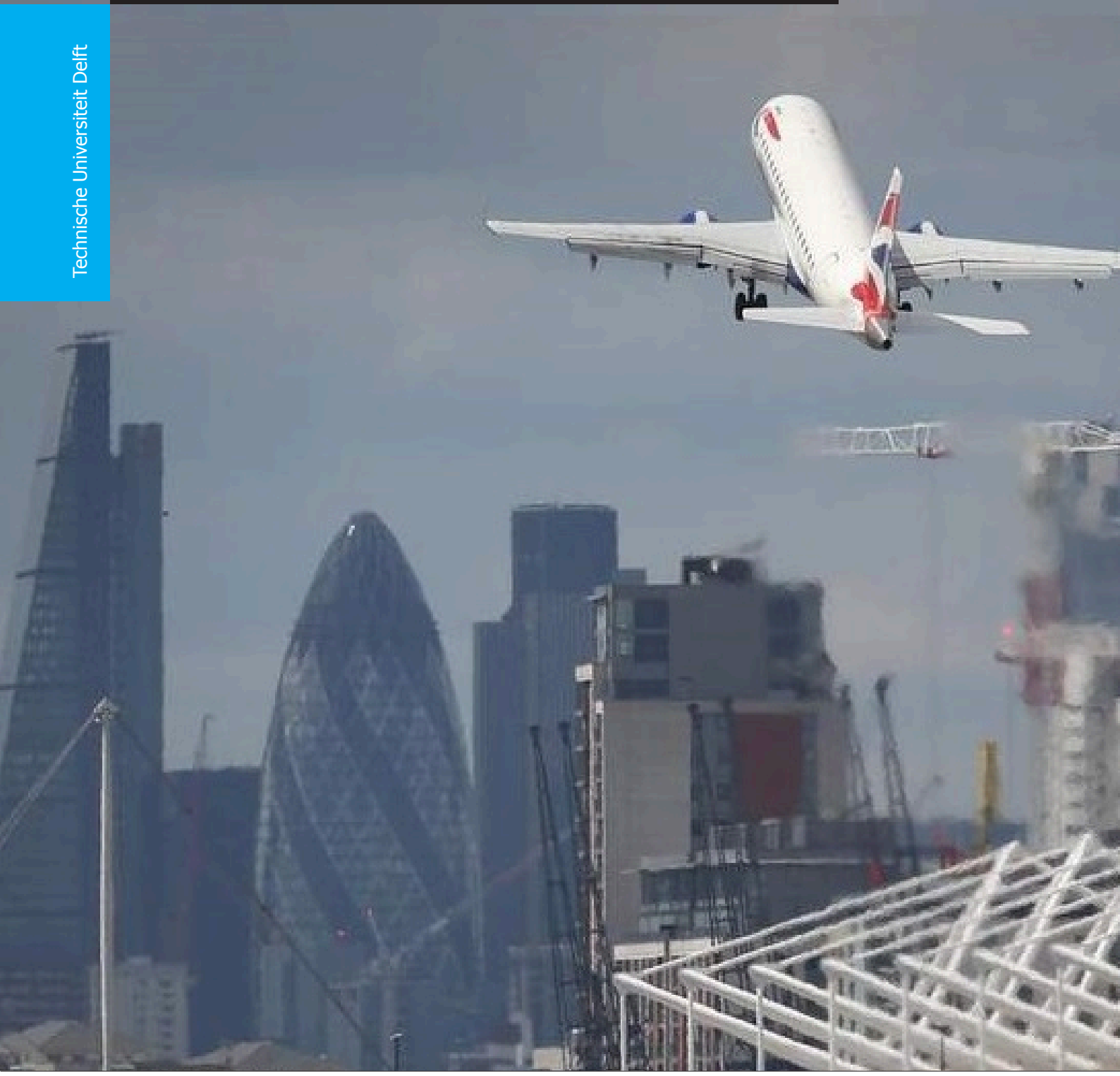
Development of an Audio Assessment Module

For Sound Engineering of Aircraft Designs

U. Mehmood

Technische Universiteit Delft



TU Delft Delft University of Technology

Challenge the future

Development of an Audio Assessment Module

For Sound Engineering of Aircraft Designs

by

U. Mehmood

in partial fulfillment of the requirements for the degree of

Master of Science
in Aerospace Engineering

at the Delft University of Technology,
to be defended publicly on Monday July 9, 2018 at 11:00 AM.

Student number:	4229630	
Supervisor:	Dr. ir. M. Snellen	
Thesis committee:	Prof. dr. D. Simons,	ANCE
	Dr. ir. M. Snellen,	ANCE
	Ir. J. Melkert,	FPP

A digital version of this thesis is available at http://repository.tudelft.nl/.

TU Delft Delft University of Technology

ACKNOWLEDGEMENTS

It has been a long journey, from searching for a graduation project to writing the acknowledgements. It is not just the end of my thesis, but also the end of my studies at the TU Delft. The past six years were challenging, amusing and enlightening at the same time. This would not have been possible without the advice and support of many people.

First and foremost, I would like to thank my supervisor Dr. ir. Mirjam Snellen for her guidance and evaluation of the work presented in here. I also wish to acknowledge the motivation and guidance provided by Dr. ir. Abhishek Sahai, who was my supervisor for the first few months of this thesis. Additionally, I would like to thank Prof. dr. Dick Simons and Ir. Joris Melkert, for being part of the assessment committee.

I would also like to thank Ana, for her valuable suggestions and her assistance during the listening tests conducted in this research. A special thanks to everyone who participated in those listening tests and were willing to listen to annoying aircraft sounds.

Last but not least, I would like to thank my parents and my siblings for their moral and emotional support in my life. I am also grateful to my other family members and friends who have supported me along the way.

U. Mehmood
Delft, June 2018

EXECUTIVE SUMMARY

One way of attaining the objective of lower resistance to aircraft noise is to use noise reduction technologies such as acoustic liners or chevron nacelles; another way of achieving this objective is to design aircraft that may inherently sound more acceptable. A first step in designing aircraft with optimal sound is the development of a module that is able to assess noise in a sophisticated way, which is the exact goal of this research.

In this research the audio assessment module (AAM) is completed with the addition of tonality, roughness and fluctuation strength. Aures' tonality metric and Daniel & Weber's roughness metric are implemented together with a fluctuation strength metric derived from the roughness metric. The AAM is now able to assess sound in terms of five sound quality metrics, specifically loudness, tonality, roughness, sharpness and fluctuation strength, but also in terms of conventional metrics such as EPNL.

The AAM is applied to a variety of sound recordings, including 255 measured aircraft flyover measurements of 26 different types of aircraft, in an attempt to find relations between design variables and the five sound quality metrics. The only significant correlations observed are those of the wingspan, wing loading and engine diameter with loudness. High values for roughness are observed for helicopter sounds due to the buzzing sound produced by the helicopter rotor, indicating that roughness is an important metric for propeller aircraft or aircraft with open-rotor engines.

Listening tests, with the aim of learning whether the psychoacoustic annoyance value (PA_{mod}) is a better annoyance predictor than EPNL, were conducted in which twenty subjects participated. The test consisted of two parts: a direct scaling part in which subjects had to give a rating between 0-100 and a two alternative forced choice part in which two sounds were presented and subjects had to indicate which one is more annoying. From the direct scaling tests it was found that the metrics did not outperform each other in terms of annoyance assessment. Both the EPNL and PA_{mod} showed a similar correlation with subjective annoyance ratings. The two alternative forced choice test was performed to investigate two sounds which contradicted each other in terms of PA_{mod} and EPNL. In some cases, the sound with the higher EPNL value was found to be more annoying, while in other cases the sound with the higher PA_{mod} value was found to be more annoying.

Some deficiencies of the PA_{mod} metric came to light during the listening tests conducted in this research. One of the reasons the PA_{mod} did not correctly predict the annoyance all the time may be the varying duration of the aircraft flyovers. For a valid comparison of two sounds in terms of PA_{mod}, the duration of the sounds has to be the same, since the values "exceeded 5% of the time" are used in the calculation of PA_{mod}. It was also found that the loudness contribution to PA_{mod} might be excessive. In addition to the loudness contribution, the effect and weighting of tonality and sharpness need further investigation. More research into PA_{mod} can potentially improve correlations with subjective evaluations. An optimized psychoacoustic annoyance model should be validated with extensive listening tests.

The design parameter analysis stated above has some limitations. The following improvements and further research possibilities for the future can be suggested. The 255 aircraft flyovers were all turbofan aircraft approaching the airport and the number of flyovers for an aircraft type ranged from 1 to 72. For future analysis, it is suggested to use a minimum of 10 aircraft flyovers for each aircraft type in order to have a more fair comparison. By taking the average of more than one aircraft flyover, the effect of disturbances such as wind will be minimized. It is also recommended to perform a design parameter analysis in a dedicated design environment using auralized sounds. By varying only one design parameter and keeping all other design parameters constant, direct relations between design parameters and sound quality metrics will become clear. Moreover, it is suggested to perform a multivariate analysis in which several design variables are examined at the same time. This gives a much more realistic picture than trying to explain the effect on a dependent variable by looking at only one single design variable.

The ultimate aim of an audio assessment module is to use it for sound engineering of aircraft designs. The psychoacoustic metrics are capable of capturing the different characteristics of sound in a more comprehensive manner than for example EPNL. Differences in sound for current aircraft or future concepts such as aircraft powered with open-rotor engines can be captured by the individual metrics. This can then be used for sound engineering of aircraft designs in which the design is modified in such a way to arrive at a sound which is as close as feasible to the target sound. In this way, it is possible to design aircraft which sound inherently more acceptable and reduce the annoyance caused to residents by aircraft noise.

CONTENTS

Nomenclature and Abbreviations

List of Abbreviations

AAM	Audio Assessment Module
ACARE	Advisory Council for Aeronautics Research
ANCE	Aircraft Noise & Climate Effects
EPNL	Effective Perceived Noise Level
ICAO	International Civil Aircraft Organization
PNL	Perceived Noise Level
PNLT	Tone-corrected Perceived Noise Level
SEL	Sound Exposure Level
SPL	Sound Pressure Level
SQ	Sound Quality

Nomenclature

ΔL	SPL excess
ΔL_m	Masking depth
Δz	Bandwidth in Bark
η_{prop}	Propulsive efficiency
A_{EK}	Amplitude of the secondary neural excitation
b	Wingspan
BPR	Engine bypass ratio
C_{FS}	Calibration factor for fluctuation strength
D_e	Engine diameter
E_{GR}	Masking intensity of broadband noise
E_{HS}	Intensity of the threshold of hearing
f	Frequency in Hz
f'	Specific fluctuation strength
f_c	Centre frequency
f_{mod}	Modulation frequency
f_{tone}	Tone frequency
FS	Fluctuation strength
FS_5	Fluctuation strength exceeded 5% of the time
H	Spectral pitch
H_m	Virtual pitch
K	Tonality
k	Crosscorrelation coefficient

K_5	Tonality exceeded 5% of the time
L	Sound Pressure Level
$L_{Aeq,T}$	Equivalent A-weighted sound level
L_{AE}	Sound exposure level
L_{BBN}	Broadband noise level
L_E	Excitation level
L_{tone}	Tone level
L_{TQ}	Level of threshold of hearing
m^*	Generalized modulation depth
N	Noisiness
$N^{'}$	Unmasked specific loudness
N_5	Loudness exceeded 5% of the time
N_{Gr}	Loudness without tones
N_{spec}	Specific loudness
N_{total}	Total loudness
PA	Psychoacoustic annoyance value
PA_{mod}	Modified psychoacoustic annoyance value
R	Roughness
r	Specific roughness
R_5	Roughness exceeded 5% of the time
ROR	Roll-off rate
S	Sharpness
S_5	Sharpness exceeded 5% of the time
T	Time period
T/W	Thrust-to-weight ratio
v	Induced pitch shift
v_j	Jet velocity
W/S	Wing loading
w_1	Bandwidth weighting function
w_2	Frequency weighting function
w_3	Excess level weighting function
w_{FR}	Fluctuation strength and roughness function in psychoacoustic annoyance model
w_{Gr}	Loudness weighting factor
w_s	Sharpness function in psychoacoustic annoyance model
w_{Ton}	Tonality function in psychoacoustic annoyance model
w_T	Tonal weighting factor
WS	Spectral pitch weights
z	Frequency in Bark

1

INTRODUCTION

A growing number of people is suffering from the noise that is being produced by aircraft due to the ever-growing airports and the continuously increasing air-traffic demand. Recently, studies have revealed that there is a direct link between aircraft noise pollution and increased risk of heart disease and stroke among residents [1], [2]. Aircraft and engine manufacturers are trying their best to create quieter aircraft designs by applying noise reduction technologies to their products. By doing this they hope to achieve lower resistance from residents to the growth of airports and air-traffic. Furthermore, the noise certification requirements enforced by the relevant governing bodies are becoming increasingly stringent. Although latest generation aircraft are 25 decibels quieter than 60 years ago [3], air traffic is increasing constantly, leading to more aircraft noise.

The focus in the aircraft industry has primarily been on noise absorbing technologies [4] and thus on making the aircraft more quiet. However, there are physical and practical limitations to these noise reduction technologies [5]. Furthermore, aircraft noise still contains adverse components such as audible tones and rapid intensity fluctuations leading to an unpleasant sound. Therefore, it is important that the problem is also tackled in a different way. The objective from this different point of view is not to make the aircraft more quiet, but to improve the sound quality such that it sounds less annoying to the residents. This involves a reduction of the noise components that might individually be contributing to the annoyance of the residents. The goal of this thesis is to implement sound quality (SQ) metrics from the field of psychoacoustics into an Audio Assessment Module (AAM), which has the future goal of being integrated in an aircraft design chain, for aircraft design with optimal sound. The three SQ metrics that are considered to be important for aircraft noise

and thus need to be implemented are: tonality, roughness and fluctuation strength. The AAM is already able to assess aircraft noise using traditional certification metrics such as Effective Perceived Noise Level (EPNL). However, the AAM should go one step beyond what currently exists for aircraft noise and assess aircraft noise using sophisticated SQ metrics. These SQ metrics can be combined to calculate a psychoacoustic annoyance value.

The structure of this report is as follows. In chapter 2 the background information is described. The problem statement is discussed and work that has already been carried out by other academics in the field of psychoacoustic and sound engineering is summarized. Chapter 3 describes the methodology used in this thesis. The validation of the implemented SQ metrics is described in chapter 4. In chapter 5 the AAM is applied to measured aircraft sounds. It is investigated how design variables have an impact on different sound characteristics. Furthermore, subjective listening tests are performed to test the capability of the psychoacoustic annoyance model. This thesis is concluded with the conclusion & recommendations in chapter 6.

2

BACKGROUND

This chapter covers the background for this master thesis. In the first section, the problem statement is discussed. In the second section, previously published literature on sound engineering and sound quality metrics is discussed.

2.1. PROBLEM STATEMENT

Besides carbon emission, noise emission has been a major concern for policy makers in the last few decades. This has led to regulations and legislation at various governmental levels which have become increasingly stringent in the last few years. The International Civil Aircraft Organization (ICAO) has defined a balanced approach to tackle the noise emission concern [6]. This implies that noise reduction should not only be achieved by optimizing operational procedures and land use planning, but also by technological improvements in the aircraft design itself. The following goal has been set by the Advisory Council for Aeronautics Research in Europe (ACARE): between 2000 and 2050, a reduction of 15 dB in EPNL, should be achieved (figure 2.1).

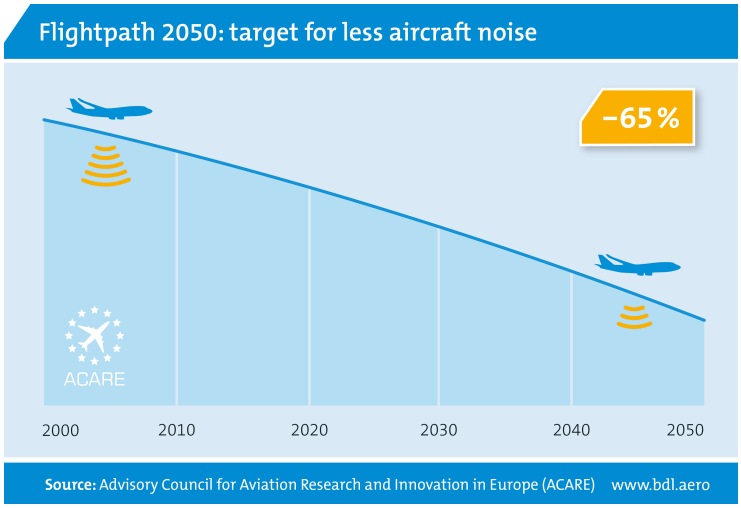


Figure 2.1: Noise emission goal set by ACARE [1]

The ICAO aids on the development and standardization of low noise operational procedures that are safe and cost-effective. The operational procedures include noise preferential runways and routes and noise abatement procedures for take-off and landing.
The appropriateness of any of these measures depends on the physical lay-out of the airport and its surroundings. These procedures have been documented in ICAO Doc 8168 [2].

Next to these global regulations, local airport rules have led aircraft manufacturers to make aircraft noise emission an important aspect in their design. Aircraft noise consists of different noise components. Major aircraft noise sources are depicted in figure 2.2. Most of the noise generated by the engine occurs due to the high velocity exhaust gases (jet noise) and the airflow in the fan (fan noise). The landing gear and high lift devices are the main contributors to airframe noise. In the last few decades, the jet noise has been reduced considerably due to the high bypass turbofan engines. High bypass ratio engines were originally developed to lower the jet velocity and thus increase the propulsive efficiency as can be seen from equation 2.1. In this equation η_{prop} is the propulsive efficiency, v_j the jet velocity and V the airspeed of the aircraft.

$$\eta_{prop} = \frac{2}{1 + \frac{v_j}{V}} \tag{2.1}$$

A positive side effect of the introduction of bypass engines was a reduction in jet noise due to the lower jet velocity, since the jet acoustic power fits a strong power law with the velocity ($\sim v_j^8$) [8]. The bypass ratio kept increasing over the years and thus the jet noise kept decreasing. Nowadays the engine fan and airframe are the dominant noise sources during

[1] Source of image: http://www.acare4europe.org
[2] https://store.icao.int/index.php/aircraft-operations-vol-i-flight-procedures-fifth-edition-2006-doc-8168-part-1-english-printed.html

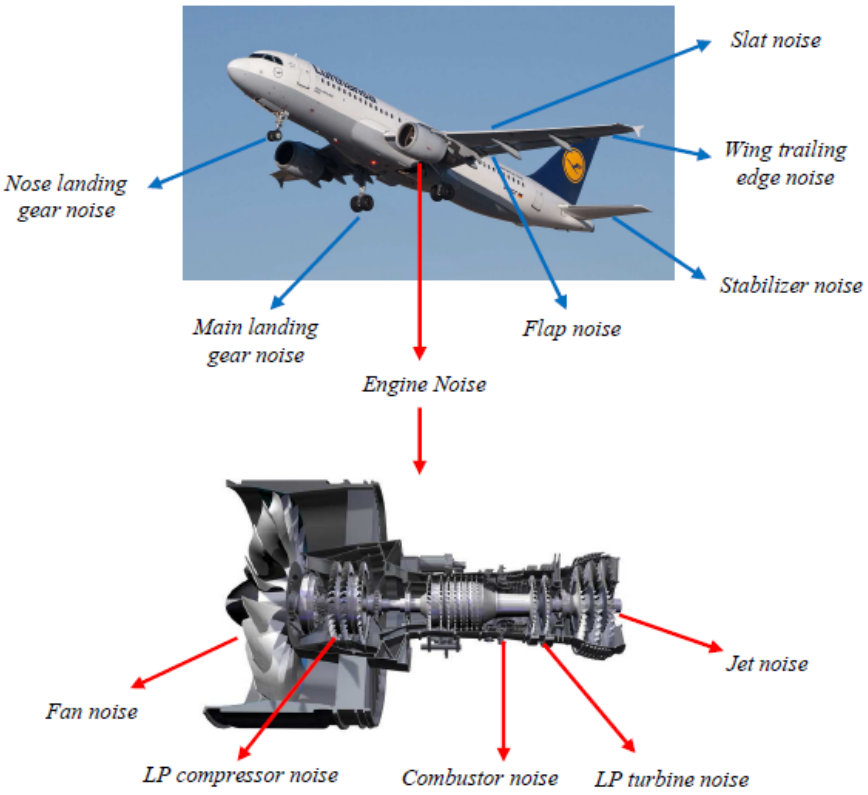


Figure 2.2: Major aircraft noise sources [7]

approach. However, jet noise remains the dominant noise source for most aircraft during take-off. Today's aircraft noise spectra differ significantly from the aircraft noise spectra of 50 years ago. This has led to a more focused approach to noise reduction. For example, chevrons, which is a sawtooth pattern on the exhaust nozzles, are being implemented on jet engines to reduce the jet noise [5]. The Bombardier CRJ900 was the first aircraft to fly with chevrons on the GE Cf34-8C5 engine. Today, there are several aircraft with chevron nacelles such as the Boeing 787. However, there is still some reluctance to systematically add chevrons since they add weight and lead to additional drag [5]. A technology which is applied more often is the acoustic liner. These liners reduce the noise within an optimized frequency range and therefore are well suited to reduce the fan noise [4].

From the examples given above, it is quite clear that the focus in the aircraft industry has primarily been on noise absorbing technologies and thus on making the aircraft more quiet. However, there are physical and practical limitations to these noise reduction technologies, such as additional weight and drag. Furthermore, aircraft noise still contains adverse components such as audible tones and rapid intensity fluctuations leading to an unpleasant sound [7]. Thus, it is necessary to not only focus on making the aircraft more quiet, but also

on improving the sound quality of the noise produced by aircraft. This can be achieved by optimizing the sound quality by analyzing any adverse sound components and removing them by altering the design. This approach is called sound engineering and is already extensively applied in industries such as automotive engineering.

For example, in the automotive industry each type of car demands a different sound. An electric car emits little noise and thus could be a hazard for pedestrians. To mitigate this hazard, car manufacturers add a synthesized sound to the car. Audi has developed e-sound, that is used on their electric e-tron models [3]. E-sound is a sound engineering technique of generating synthetic engine sounds for the essentially silent vehicle in order to warn pedestrians and cyclists of the cars proximity. On the other hand, car manufacturers are using sound engineering to intensify the engine-sound of their cars. Achieving an optimal sound can be done by altering the design, as is done on the Lexus LFA [4]. Components were designed on the engine which directed the engine sound towards the driver.

The sound engineering approach described above can potentially also be adopted by the aircraft industry to improve the sound quality of aircraft. This will improve the acceptance of aircraft noise and lead to lower resistance of residents to the construction of airports and air-traffic expansion. For that reason, an AAM is currently being developed by the Aircraft Noise & Climate Effects (ANCE) section at the TU Delft. The AAM will serve the more overall goal of being able to design aircraft with optimized sound by integrating the module into the aircraft design chain. Such a design chain is depicted in figure 2.3.

Figure 2.3: Methodology for low-annoyance aircraft design [9]

After the initial aircraft design, the flight path is calculated and source noise is predicted. The next step in the design chain is the auralization (i.e. aural simulation) of aircraft sound. This auralized aircraft noise can then be used for annoyance assessment and the results can subsequently be fed back to the aircraft design tool. To the best of the authors knowledge, there is no noise assessment module which is integrated into the aircraft design chain. A noise assessment module integrated in the design chain can help to identify negative sound effects at an early stage, reducing the need to apply expensive noise reduction technologies

[3] https://www.designboom.com/technology/audi-e-sound-for-electric-cars/
[4] http://blog.lexus.co.uk/a-closer-look-inside-the-lfas-v10/

such as chevron nacelles at a later stage. In addition, by using the AAM for the noise produced by current aircraft, insights can be obtained in the most annoying aircraft and its components. This can be used to initiate the process described by figure 2.3.

To be able to design aircraft with optimal sound, suitable noise assessment metrics are needed. The conventional metrics used for noise assessment can be divided in three main categories: Loudness based weighted Sound Pressure Level (SPL) metrics, annoyance based Perceived Noise Level (PNL) metrics and multiple event average level metrics. All these metrics are briefly explained below.

Loudness based weighted SPL metrics

SPL is a measure of sound pressure and uses a logarithmic scale to represent the sound pressure of a sound relative to a reference pressure. The reference sound pressure is typically the threshold of human hearing ($2{\cdot}10^{-5}Pa$). Loudness is the subjective perception of sound pressure. This should not be confused with objective measures of sound pressure such as SPL. Loudness is defined as the auditory sensation in terms of which sound can be ordered on a scale from quiet to loud [10]. Loudness can be expressed in phon. A phon is defined as follows: a sound has a loudness level of X phons if it is equally loud as a tone with a SPL of X dB at 1 kHz. A more detailed explanation of loudness is given in section 2.2. The three loudness based weighted SPL metrics are the following.

- A-weighted SPL: Non-linear frequency selectivity of the ear plays an important role in loudness metrics. This can be seen in the equal loudness contour in figure 2.4.

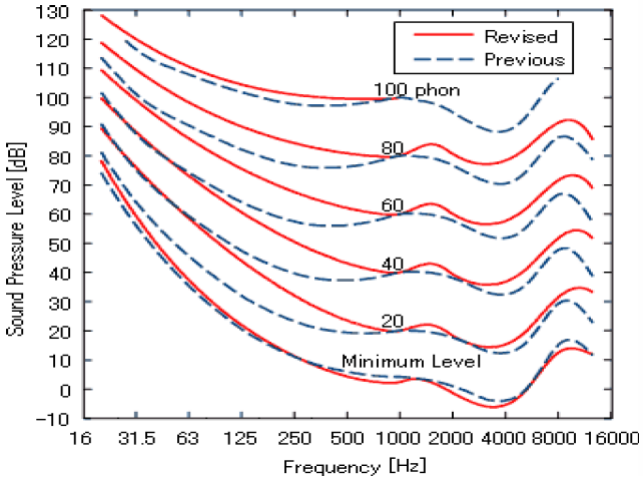


Figure 2.4: Equal loudness contours [5]

[5]Source of image: http://en.wikiaudio.org

The A-weighted metric was developed to approximate the fact that the human hearing system perceives high frequency sound as louder than low frequency sound of the same SPL. When assessing the noise impact, higher frequencies are more highly weighted than low frequencies. In the equal loudness contours, the 40 phon curve was chosen as a reference for this metric.

- C-weighted SPL: The difference with A-weighted SPL is that this metric was developed to approximately follow the 100 phon equal loudness contour. This means that this metric is for the assessment of louder noise levels than the A-weighted SPL (40 phon).

- Sound Exposure Level (SEL): In this calculation, the noise impact from the A-weighted SPL is integrated over time to account for the duration. The SEL value thus corresponds to the constant sound level which has the same amount of acoustic energy in 1 second as the whole noise event.

Annoyance based Perceived Noise Level

The annoyance aspect of noise is not only dependent on the loudness. The duration of the noise and the presence of tones are other important factors to determine the annoyance caused by a sound signal. The equal loudness contours given above are not derived for broadband noise, such as aircraft noise. Therefore, equal noisiness curves were introduced (figure 2.5). Noisiness is expressed in noy N.

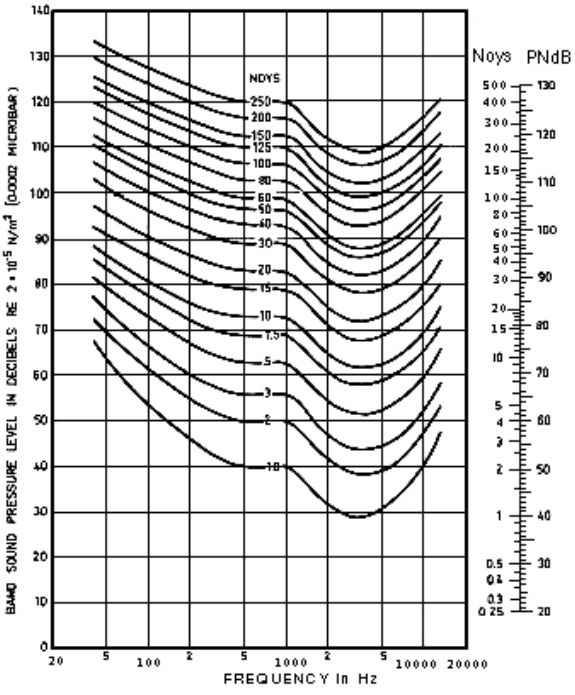


Figure 2.5: Equal noisiness contours [6]

[6] Source of image: https://www.sfu.ca/sonic-studio/handbook/Perceived_Noise_Level.html

- Perceived Noise Level (PNL): This metric was developed in the 1960s and had the goal to provide a means of certifying aircraft noise. The equal noisiness curves are used to calculate the PNL in PNdB using equation 2.2.

$$PNL = 40 + 33.3log(N) \tag{2.2}$$

- Tone-corrected Perceived Noise Level (PNLT): It was found that the presence of discrete tones in broadband noise make the sound more annoying. Therefore, a tone correction [11] is added to the PNL value.
- Effective Perceived Noise Level (EPNL): The final addition to the PNL metric was the inclusion of a duration correction [11] that would ultimately yield a single value measure for the annoyance caused by aircraft noise for a single event.

Multiple event average level metrics

Single event noise metrics such as EPNL are most commonly used for aircraft noise certification. However, multiple event metrics are also convenient for community noise assessment. These metrics are intended to focus on how residents are affected by multiple aircraft flyovers in a specific time period.

- Equivalent A-weighted sound level for multiple events: The SEL values specified above are for a single event. When summed energetically for a time period, an equivalent sound level can be calculated for that period of interest. This can be done for multiple events in the following way:

$$L_{Aeq,T} = 10\ log_{10} \sum_{j=1}^{n} 10^{L_{AE}(j)/10} - 10\ log_{10}\ T \tag{2.3}$$

where n is the number of events, T is the time period and $L_{AE}(j)$ is the SEL value for the j^{th} event.
- Day-night average level: In this metric a penalty is given during night times, since noise is more annoying when the background or ambient noise is low.
- Day-evening-night average level: This is a variation of the previous metric, since in this case a penalty is also given for events in the evening. The evening penalty is lower than the penalty for events in the night.

The metrics described above are quite basic metrics which attempt to assess the noise annoyance experienced by residents. The loudness based metrics do not take into account effects of tones. Furthermore, the approximation to a 40 phon (A-weighting) or a 100 phon (C-weighting) equal loudness curve is not really representative of the loudness experienced

in the vicinity of an airport. The tone-correction and duration correction in the annoyance based PNL metrics are also quite basic. For example, for the tone correction, the frequency spectrum is divided in a very broad and coarse way. Furthermore, only the single most protruding tone is taken into account, neglecting the other tones which could still affect the annoyance perception. The penalties given in the multiple event average level metrics are rather arbitrary. Also, tonal content or the effect of frequency is not taken into account. In general, all these metrics were developed over 50 years ago and since then much more detailed knowledge about noise annoyance experienced by humans has been accumulated.

The AAM is already able to assess aircraft noise in the traditional certification metrics such as SEL and EPNL. However, the AAM is supposed to go one step beyond what currently exists for aircraft noise assessment. Models which reflect the physics behind the noise and take into consideration the way in which the human hearing system processes sounds are needed. Psychoacoustics is the scientific field which attempts to bridge the gap between physical and subjective evaluations of sound. In the field of psychoacoustics a lot of research has been conducted on how sound is perceived by the human ear. This has led to much more sophisticated measures of perceived sound based on the characteristics of the sound. Sound attributes such as loudness, tonality, sharpness, roughness and fluctuation strength all contribute to noise annoyance [10],[12]. A spectrogram is a visual representation of audio. A spectrogram for an aircraft flyover is given in figure 2.6.

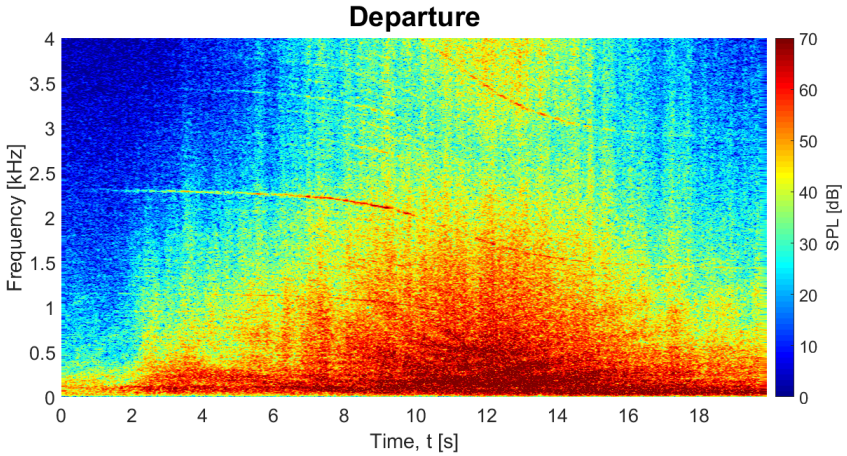


Figure 2.6: Departure spectrogram of a Boeing 737-700

High lift devices, together with landing gears, are the main sources of airframe noise during departure and approach. Airframe noise is mainly broadband noise and thus contains energy over a wide range of frequencies. In the sample spectrogram it can be seen that when the aircraft is overhead, the energy is spread over a wide range of frequencies. Loudness is the subjective perception of the magnitude of a sound. When a spectrogram contains more

spectral energy, the loudness value will be higher. However, it must be noted that loudness is not only dependent on the SPL, but also on frequency and duration.

The lines in the spectrogram indicate tones. The lines change in frequency over time due to the Doppler shift. The most prominent tone is the tone generated by the engine fan and is called the Blade Passage Frequency (BPF) tone [13]. The BPF can be calculated with equation 2.4, where B is the number of blades and n is the rotational velocity in rpm. Higher harmonics of the BPF are also present [14].

$$BPF = \frac{B \cdot n}{60} \tag{2.4}$$

During take-off, buzzsaw tones can also be heard, which are generated when the speed at the blade tips is supersonic. These buzzsaw tones are a harmonic series of tones with a low fundamental frequency [13]. All these tones should be captured with a suitable tonality metric.

Sharpness deals with the high frequency content of a sound. Applying it to a spectrogram, a sound will have a higher sharpness when there is a lot of spectral energy concentrated at the higher frequencies (above 3000 Hz). Roughness deals with fast fluctuations in sound, while fluctuation strength deals with slow fluctuations in sound. An important source of roughness in aircraft sound is a propeller engine. Fluctuation strength plays a negligible role in the assessment of aircraft noise, but is still implemented since it is very similar to roughness.

Algorithms have been developed that take into account the aforementioned sound attributes in order to improve the sound quality of products. In sound engineering, these models are applied to products to improve sound quality. The development of a noise assessment module, which is the main objective of the thesis, forms the foundation for sound engineering of aircraft designs. In the AAM, loudness and sharpness have already been incorporated. That is why the focus of this master thesis will be on the tonality, roughness and fluctuation strength SQ metrics.

2.2. ACADEMIC LITERATURE

In this section the academic literature on the psychoacoustic metrics will be presented. First of all, the concepts which play an important role in psychoacoustics will be discussed. The concepts explained are the following three: non-linear frequency sensitivity, masking and critical bands.

Frequency sensitivity

The frequency sensitivity of the human ear is captured by the equal loudness contours, which were depicted in figure 2.4. The most important conclusion that can be drawn from the equal loudness contours is that an increase in level at low frequencies will be perceived as a higher change in loudness, than at high frequencies with the same level increase. Applied to aircraft noise, this means that for example fan tones generally occurring around 1 kHz and above could have an increased SPL and still be perceived as loud as before, while low frequency jet noise would be perceived as clearly louder with the same increase in SPL [7].

Masking

Masking refers to the decrease in audibility of a sound due to the presence of another sound. The effect of masking is measured quantitatively by determining the masked threshold, which is the SPL of a test tone (usually sine wave) necessary to just be audible in the presence of a masker (other sound). In general, high frequency sounds will be masked by low frequency sounds if both are present simultaneously. A tone or narrowband noise will mask frequencies above the center frequency more effectively than frequencies below the center frequency. This is called the upward spread of masking. Masking of a test tone by a masker is illustrated in figure 2.7.

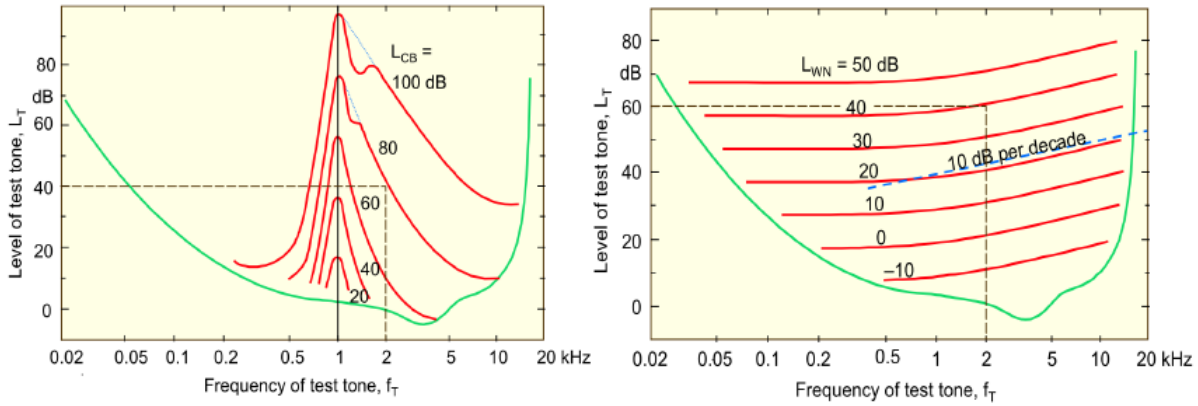


Figure 2.7: Masking of 2 kHz test tone by narrowband noise centered at 1 kHz (left) and white uniform noise (right). The green line indicates the threshold of hearing. CB = critical band and WN = white noise. [15]

In the left plot, it can be observed that the narrowband noise will mask the test tone of 2 kHz and 40 dB if it has a SPL of 80 dB or above. Also, it can be observed that the masking is more effective towards higher frequencies (upward spread). Each tone or narrowband noise will have its own masking pattern. In the right plot of figure 2.7, the masking due to white uniform noise can be seen. For all dB levels of white noise the following holds: the

curve is constant until approximately 500 Hz and has a 10 dB per decade slope after 500 Hz. To mask the test tone of 2 kHz and 40 dB, a level of 30 dB of the white noise is already sufficient. This is much lower than the 80 dB found for the narrowband noise. This observation implies that broadband noise such as jet noise is very effective in masking other tones. Masking effects can be measured not only when masker and test sound are presented simultaneously, but also when they are not presented simultaneously. The test sound can also be presented before the masker stimulus is switched on or after termination of masker stimulus. The masking effect produced in the latter case is referred to as postmasking, while in the first case it is called premasking [10]. Premasking, simultaneous masking and postmasking are schematically drawn in figure 2.8.

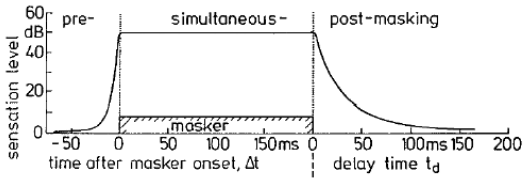


Figure 2.8: Schematic drawing to illustrate and characterize the regions within which premasking, simultaneous masking and postmasking occur [10].

Critical bands

The third important concept is the division of the audible frequency range of the human ear into frequency bands. These frequency bands are called the critical bands. Thus, the critical bandwidth corresponds to the frequency resolution of the ear. The critical band rates z are measured in Bark and have a range from 0 to 24. The critical bandwidth as a function of frequency is given in figure 2.9 and can be computed using equation 2.5.

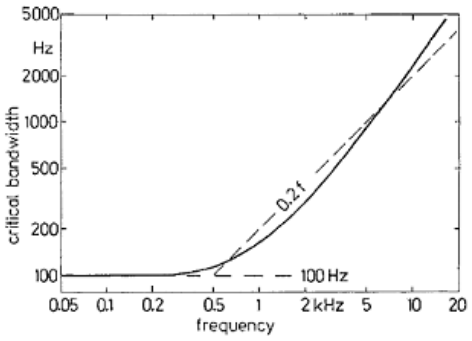


Figure 2.9: Critical bandwidth approximation. For frequencies until 500 Hz the critical bandwidth is estimated to be 100 Hz, and for higher frequencies it is estimated to be $0.2 \cdot f$ [10]

$$z = 13 \tan^{-1}\left(0.76 \frac{f}{1000}\right) + 3.5 \tan^{-1}\left(\frac{f}{7500}\right)^2 \tag{2.5}$$

An important implication of the critical band concept is that two tones present within the same critical band will not be heard as two separate tones and the sounds will affect each others perception due to masking. Furthermore, loudness stays the same when increasing the bandwidth of narrowband noise up to the critical bandwidth while keeping the SPL constant (but the overall sound pressure level increases). If the narrowband noise bandwidth exceeds the critical bandwidth, the loudness increases while keeping the SPL constant.

2.2.1. SPECIFIC AND TOTAL LOUDNESS

Loudness is the subjective perception of the magnitude of a sound. It depends on the frequency (figure 2.4), intensity and the duration. Loudness is given in phon when on a logarithmic scale, and in sone when using a linear scale. Loudness has been standardized in ISO 532-1 [16].

First, the SPL of the sound is calculated and transformed to 1/3 octave band levels. These 1/3 octave band levels are then transformed to a SPL value in each critical band (critical band level). This is done using the equal loudness contours. This is followed by the calculation of the excitation pattern, which is analogous to the amount of neural activity in the human ear [10]. The excitation pattern is calculated as follows. The excitation level of the component is set to the SPL of the component at the place on the Bark scale which corresponds to the frequency of the component. As proposed by Terhardt [17], the excitation decreases with a constant slope towards lower frequencies and with a level dependent slope towards higher frequencies. In psychoacoustics, the excitation pattern is commonly used to model auditory masking [18]. Subsequently, the specific loudness, which is the loudness in each critical band expressed in sone/Bark, is calculated with the following equation:

$$N_{spec}(z) = 0.0635 \cdot 10^{0.025 L_{TQ}(z)} \left[(0.75 + 0.25 \cdot 10^{0.1 (L_E(z) - L_{TQ}(z))})^{0.25} - 1 \right] \tag{2.6}$$

where L_E is the excitation level and L_{TQ} is the threshold in quiet. In the next step, it is checked whether the specific loudness in each critical band is either partly or completely masked by the so-called accessory loudness that is caused by the excitation over the current critical band from a sound concentrated on another critical band. The unmasked specific loudness $N^{'}$ in each critical band is then determined and used to compute the total loudness N_{total} according to equation 2.7.

$$N_{total} = \int_0^{24} N^{'}(z)\, dz \tag{2.7}$$

Since aircraft noise is not steady but time-dependent, the time-varying loudness calculation is very relevant. The model of loudness for time-varying sounds takes into account

temporal masking. Postmasking is the dominant temporal masking effect and premasking plays a relatively secondary role [10], therefore only postmasking is taken into account in Zwickers model [19].

Figure 2.10 shows an example of how loudness versus time is established for a 5 kHz tone burst, with a duration of 10 ms (dotted line) or 100 ms (solid line). The total loudness corresponds to the area under the unmasked specific loudness pattern. Hence, the curve of total loudness is lower for the burst with a duration of 10 ms.

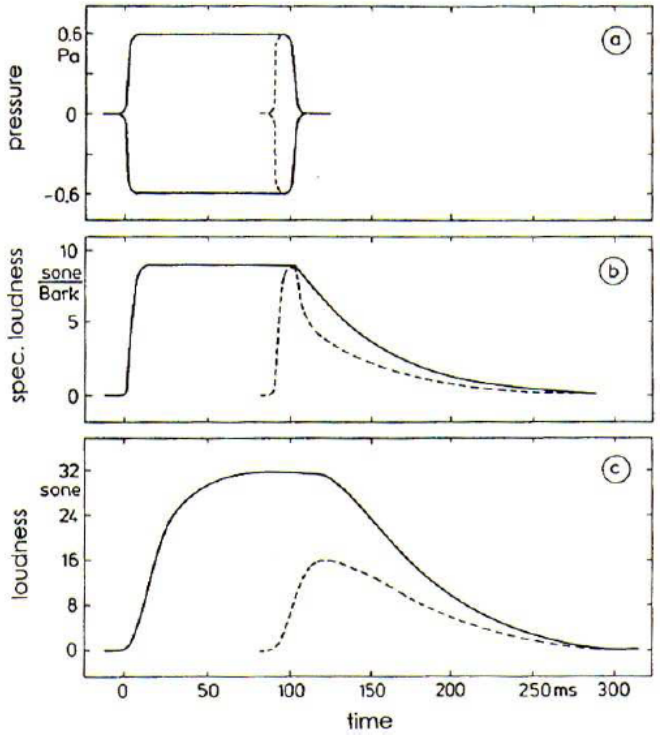


Figure 2.10: Establishment of the loudness of a 5 kHz tone burst with duration of 100 ms (solid line) and 10 ms (dotted line) (a) temporal envelope of both tones (b) specific loudness in the 19th critical band. (c) loudness as a function of time [10]

Although Zwickers model allows the calculation of loudness for time-varying sounds, it does not give one overall loudness value for the sound. However, for many sounds where loudness varies with time, it has been found that the loudness exceeded 5% of the time (N_5) is a reasonably good measure of perceived loudness [16]. As an example, the timevarying loudness is plotted for an aircraft flyover in figure 2.11. As can be seen, the signal duration is 15 seconds, and thus the N_5 value, is the value exceeded for 0.75 seconds. This value is indicated by the dashed line in the figure.

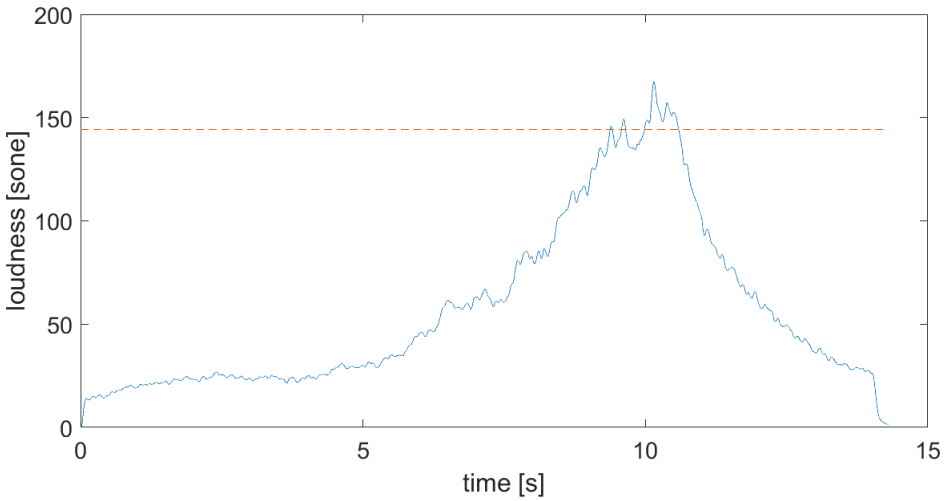


Figure 2.11: Timevarying loudness for an aircraft flyover

2.2.2. SHARPNESS

Sharpness is a measure of the high-frequency content of a sound. A sound with more high-frequency content is perceived to be sharper. The unit of sharpness is acum and one acum is referenced to a narrowband noise one critical band wide, with a center frequency of 1 kHz and having a SPL of 60 dB. Many methods exist which attempt to capture the sharpness of a sound, such as the method of Zwicker and Fastl [10], Aures' sharpness method [20] and the method of von Bismarck [21]. Sharpness in the standard DIN 45692 [22] is similar to the sharpness method of von Bismarck. Sharpness according to the method of von Bismarck can be calculated with the following equation:

$$S = c \times \frac{\int_0^{24} N'(z) g(z) z \, dz}{N} \tag{2.8}$$

where c is a constant depending on the normalization of the reference sound, N' is the specific loudness at critical band rate z and $g(z)$ is a weighting function. The weighting function is given in equation 2.9. For high critical band rates, it causes a strong increase in sharpness.

$$g(z) = \begin{Bmatrix} 1 & z \leq 16 \\ 0.066 e^{0.171 z} & z > 16 \end{Bmatrix} \tag{2.9}$$

2.2.3. TONALITY

This sound quality metric is concerned with the tonal prominence of a sound. Terhardt's tonality metric [23] forms the theoretical basis of tonality metric definition. Developed in 1982, it serves as the basis for many other tonality metrics. This method is based on the

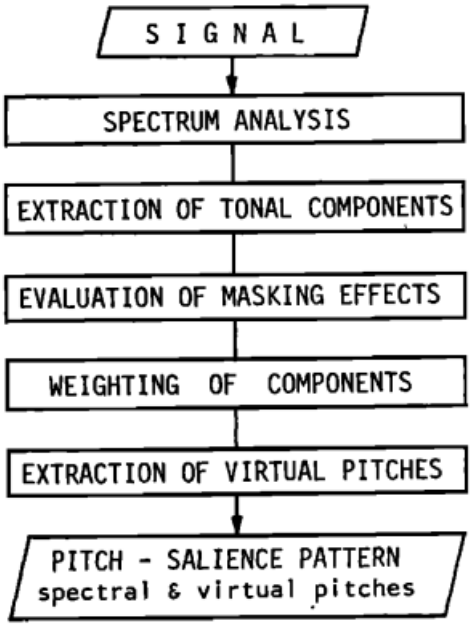


Figure 2.12: Overview of Terhardts tonality metric [23]

virtual pitch theory. According to the virtual pitch theory, the first six to eight harmonics of a complex tone can be perceived as separate spectral pitches. From these spectral pitches, the virtual pitches can be deducted. A schematic overview of Terhardt's tonality method is given in figure 2.12.

In the study of Zhang & Shreshta [15], this algorithm is implemented in Matlab and in VCC+. First of all, the signal is transformed to the frequency domain using a Fast Fourier Transform (FFT). This is followed by the extraction of the tonal components. The spectral samples are investigated and it is tested whether the SPL of the i^{th} sample is higher than the next lower $(i-1)$ sample and the next higher $(i+1)$ sample. It is then further tested whether a candidate found by this criterion meets the condition specified in equation 2.10.

$$L_i - L_{i+m} \geq 7\ dB\ for\ m = -3,-2,..,2,3 \tag{2.10}$$

If this condition is met, it is assumed that the group of seven spectral samples $(i-3, i-2, .., i+2, i+3)$ forms a tonal component. The value of 7 dB is determined empirically.

The center frequency of the tonal component is not equal to the frequency of the i^{th} sample, but is calculated using equation 2.11.

$$f_c = f_i + 0.46 \cdot (L_{i+1} - L_{i-1}) \tag{2.11}$$

The next step in the method is the evaluation of masking effects. The SPL excess is calculated to determine the aurally relevant tones. The SPL excess ΔL can be calculated using equation 2.12, where n is the number of tonal components.

$$\Delta L_i = L_i - 10log_{10}\left\{\left[\sum_{k\neq i}^{n} A_{Ek}(f_i)\right]^2 + E_{Gr}(f_i) + E_{HS}(f_i)\right\} dB. \tag{2.12}$$

The term A_{Ek} is the amplitude of the secondary neural excitation at frequency f_i due to a k^{th} tonal component and can be calculated with equations 2.13-2.15.

$$A_{Ek}(f_i) = 10^{\frac{L_{Ek}(f_i)}{20}} \tag{2.13}$$

$$L_{Ek}(f_i) = L_k - s(z_k - z_i) \tag{2.14}$$

$$s = \begin{Bmatrix} 27 & f_i \leq f_k \\ -24 - \frac{230}{f_k} + 0.2L_k & f_i \geq f_k \end{Bmatrix} \tag{2.15}$$

The second term in the SPL excess equation, E_{Gr} refers to the masking intensity of the broadband noise encompassing the tone. It is equal to the sum of the broadband noise intensities surrounding the tone one Bark around it. The final term, E_{HS} is the intensity at the threshold of hearing and can be calculated using equation 2.16.

$$E_{HS}(f_i) = 3.64\frac{f_i}{1000}^{-0.8} - 6.5e^{-0.6(\frac{f_i}{1000}-3.3)^2} + 10^{-3}\left(\frac{f_i}{1000}\right)^4 dB \tag{2.16}$$

This calculation is done for all tonal components from $i = 1$ to $i = n$. Positive values for ΔL_i indicate components which are aurally relevant.

Due to interaction of simultaneous spectral components in the auditory system, the spectral pitches are not exactly identical to the pitches of isolated tones with the same frequency. The individual spectral pitch of a tonal component expressed in pitch units (pu) can be calculated using equation 2.17, where v_i is the induced pitch shift which can be determined using equation 2.18

$$H_i = f_i(1 + v_i) \tag{2.17}$$

$$v_i = 2\cdot 10^{-4}\Big(L_i - 60\Big)\Big(f_i - 2\Big) + 1.5\cdot 10^{-2}e^{\frac{-\Delta L_i^1}{20}}\Big(3 - ln(f_i)\Big) + 3\cdot 10^{-2}e^{\frac{-\Delta L_i^2}{20}}\Big(0.36 + ln(f_i)\Big) \tag{2.18}$$

where, ΔL_i^1 and ΔL_i^2 are given by equations 2.19 and 2.20.

$$\Delta L_i^1 = L_i - 20log_{10}\sum_{k=1}^{i-1} 10^{L_{Ek}(f_i/20)} \tag{2.19}$$

$$\Delta L_i^2 = L_i - 20log_{10}\sum_{k=i+1}^{n} 10^{L_{Ek}(f_i/20)} \tag{2.20}$$

The true spectral pitch thus takes into account the pitch shifts, whereas for the nominal spectral pitch ($H = f_i$) the pitch shifts are ignored. The next step in Terhardt's method is the weighting of components. Each tonal component is weighted based on its SPL excess level and frequency. The weighting function is given in equation 2.21.

$$WS_i = \left[1 - e^{\frac{\Delta L_i}{15}}\right]\left[1 + 0.07\left(\frac{f_i}{0.7} - \frac{0.7}{f_i}\right)^2\right]^{-0.5} \tag{2.21}$$

The spectral pitch weights WS_i versus spectral pitch is called the spectral pitch pattern. This serves as the basis to calculate the virtual pitch, which is the final step in the method. As stated earlier, it is one of the main principles of virtual pitch theory that virtual pitch, even though it is a fundamentally different perceptual attribute, is dependent on the pitches of the dominant spectral components. The nominal virtual pitch can be calculated using equation 2.22, while the virtual pitch accounting for pitch shifts and interval stretch (true virtual pitch) can be calculated using equation 2.23.

$$H_{im} = \frac{f_i}{m} \tag{2.22}$$

$$H_{im} = \frac{f_i}{m}\left(1 + v_i - sign(m-1)10^{-3}\{18 + 2.5m - (50 - 7m)\frac{f_i}{m} + 0.1(\frac{f_i}{m})^{-2}\}\right) \tag{2.23}$$

The virtual pitch is represented by the m^{th} subharmonic of the i^{th} relevant component. Also, a weight is given to these virtual pitches depending on on how well the determinant component fits into a harmonic series with other components. Furthermore, a correction is applied to the weight for the fact that virtual pitch only exists in the human frequency range. The virtual pitch value (H_{im}) and its corresponding weight and frequency form the virtual pitch pattern.

An example is illustrated in figure 2.13. It represents a pitch analysis of an artificial complex tone from the study of Zhang & Shreshta [15].

Five plots can be seen in the figure. In the most upper plot the SPL versus frequency is plotted. The second plot depicts the tonal components extracted by the algorithm according to equations 2.10 and 2.11. It can be observed that the tonal components correspond to the peaks which can be seen in the first plot. The third plot shows the tonal components which have an SPL excess (equation 2.12) larger than zero. Only six of the nine components have an SPL excess larger than zero. In the fourth plot the spectral pitch weight (equation 2.21) is plotted. The spectral pitch weight is the highest for the tonal component at 500 Hz. The virtual pitches (equation 2.23) are subharmonics of the six spectral pitches indicated

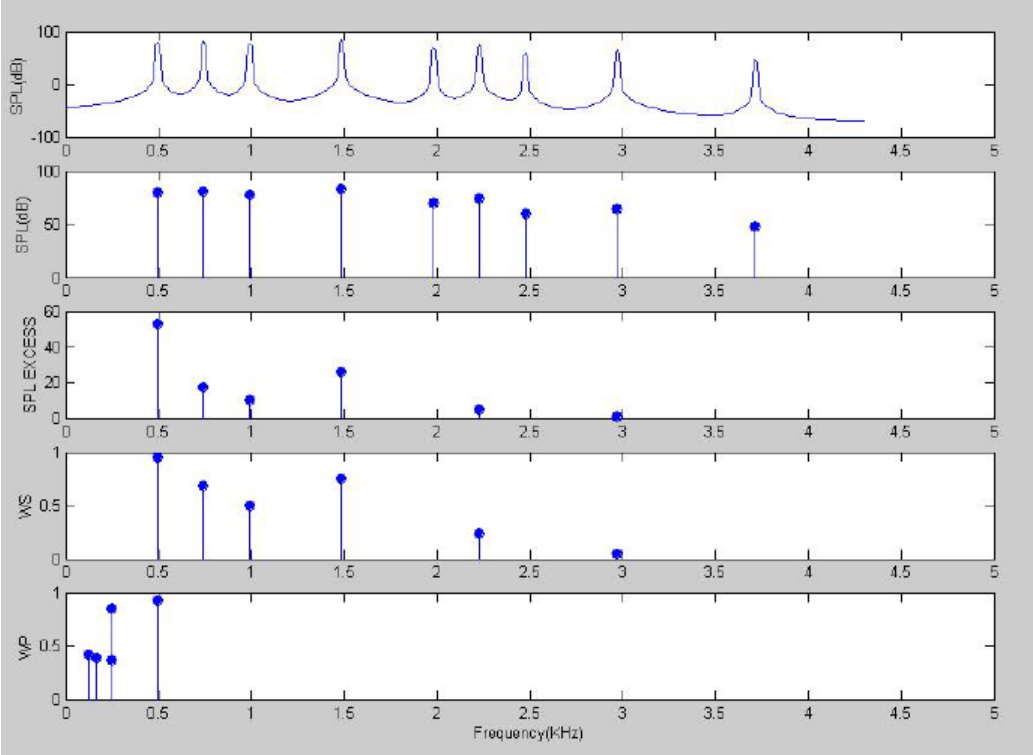


Figure 2.13: Pitch analysis of complex tone [15]

in the fourth plot. Only the five most prominent virtual pitches are shown in the last plot in order to simplify the plot. It can be observed that the most prominent virtual pitch is the one which corresponds to the fundamental frequency of 500 Hz. The fourth and fifth plot are the important outputs for this tonality metric. This metric does not calculate a single number for tonality.

Due to the assumptions Terhardt made while developing the method, equation 2.23 can only be applied when the pitch value is lower than 500 pitch units (pu). Due to the simplicity of the equation, virtual pitches higher than 500 pu may deviate considerably from the truly perceived pitch. As stated above, another drawback of this method is that it does not give a single number describing the tonality perceived by the human ear. Instead of a spectral pitch and virtual pitch pattern, a single number describing the tonality of a sound perceived by the human hearing system would be more useful. A recommendation is made by Zhang and Shrestha to find a weighting function which can be applied to both the spectral and virtual pitch patterns and extract a single number from it for tonality. Furthermore, it is reasoned by Zhang and Shrestha that the precision of the tonality calculation will increase when a proper procedure can be found that introduces loudness into tonality. Both of these drawbacks are eliminated when using Aures' tonality metric [24]. Aures found that if loudness is introduced into the tonality metric, it will have better correlation with sub-

jective results and a more standardized evaluation of masking effects. Aures' Tonality metric has been widely used in the automotive industry to improve the sound quality of cars [25],[26]. Furthermore, Angerer et al. [27] and More [12] found that the annoyance due to aircraft tonal noise could be predicted with a reasonably high correlation by using Aures' tonality metric.

An examination of Aures' tonality metric has been conducted by Hastings and Davies [28]. Sounds with different tonal content were compared by subjects to assess the influence of bandwidth and frequency on tonality. Two types of sounds were used: tones and bandpass filtered random noise. For the filtered noise tests, the frequency response of the filter was varied by changing the location of the half power points and the roll-off rate. To some extent, the dependency on frequency and bandwidth of tonality were confirmed by the tests. However, sounds generated with different order filters did not result in equal perceived tonality.

This can also be observed in figure 2.14. In this figure the estimated tonality of the bandwidth and filter order test is plotted, as a function of percent of critical bandwidth. The subjective response ratings are obtained using listening tests in which 26 subjects participated. The blue and red line correspond to the filter order tests. The solid line indicates weighting function w_1, which captures the bandwidth effect in Aures' model. It can be seen that there is a significant difference between the curves for the fourth order filter and sixth order filter.

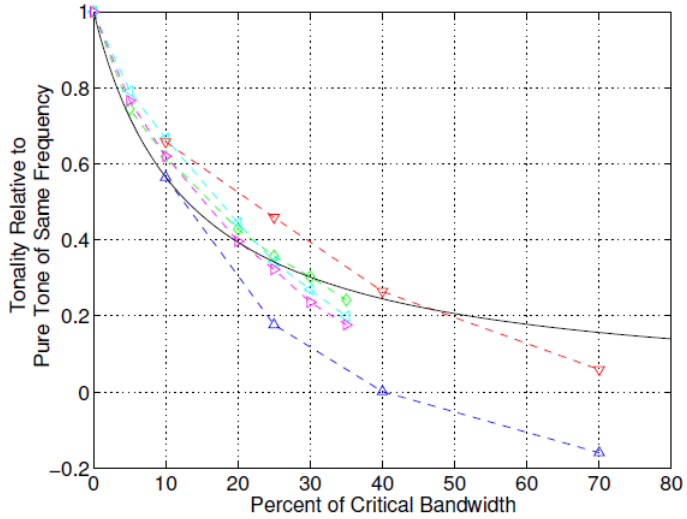


Figure 2.14: A comparison of the subjective response ratings for the bandwidth and filter order tests. Symbols: (△) 4^{th} order narrowband noise at 13.5 Bark, (▽) 6^{th} order narrowband noise at 13.5 Bark, (◇) narrowband noise at 4.5 Bark, (◁) narrowband noise at 9.5 Bark, (▷) narrowband noise at 12.5 Bark, (-) weighting function w_1 [28]

According to Hasting and Davies, this may be an indication that the half power points are not good measures of the perceived bandwidth. In another study by Hastings et al. [29] it was found that roll-off rate of the tonal component affected the tonality perception.

Besides the more sophisticated tonality metrics such as Terhardt's tonality and Aures' tonality, there exist some metrics which quantify the tonal content of a sound in a more basic manner. Examples of such metrics are the Tone-to-Noise Ratio, Prominence Ratio and the Joint Nordic Method. The Tone-to-Noise Ratio is defined as the ratio of power contained in the tonal component under investigation to the power contained in the critical band centered around that tone, but not including the part of the spectrum that includes the tonal feature. This is schematically depicted in figure 2.15. The primary tone is the tone with the highest level in the band. The tone with the second highest level in the critical band centered on the primary tone, is called the secondary tone. If the primary and secondary tone are close to each other in frequency, their tone powers are combined.

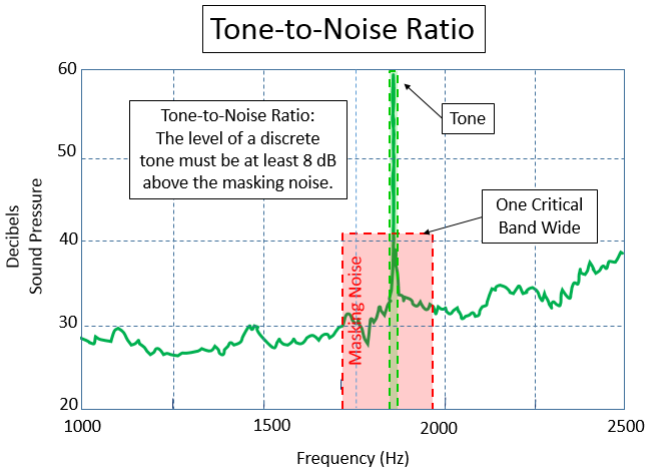


Figure 2.15: Schematic illustration of the Tone-to-Noise Ratio calculation [7]

The Prominence Ratio is defined as the ratio of the power contained in the critical band centered on the tone under investigation to the average power contained in the two adjacent critical bands, one below, and one above (see figure 2.16). When multiple tones are present in a critical band, the Prominence Ratio is more effective than the Tone-to-Noise ratio. However, in the case of multiple tones in adjacent critical bands, the Tone-to-Noise ratio is more effective.

In the Joint Nordic Method [30] the sound pressure levels of the tones are determined from

[7] Source of image: www.community.plm.automation.siemens.com/t5/Testing-Knowledge-Base/ Tone-to-Noise-Ratio-and-Prominence-Ratio/ta-p/423296

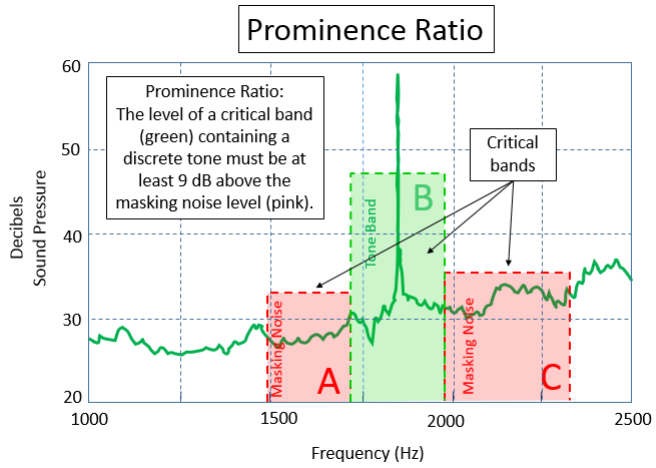


Figure 2.16: Schematic illustration of the Prominence Ratio calculation [7]

a narrow-band power spectrum. Tones in the same critical band are added to give the total tone level. Then the sound pressure level of the masking noise is determined, using the average noise level in that critical band. The total tone level and masking noise level is used to calculate the Tonal Audibility. The Tonal Audibility is used to penalize the measured average A-weighted sound pressure level.

Since the Tonal Audibility is used to penalize the A-weighted SPL, the Joint Nordic Method has the same limitations as the conventional A-weighting metric. Furthermore, the penalties given are rather arbitrary [7]. Tone-to-Noise ratio and Prominence Ratio are widely used for evaluating tonality of IT devices. In the paper of Sottek [31] it is argued that both these metrics do not correspond well or even at all to the tonality caused by narrowband noise or complex tones.

In another study [13], four different tonality indicators (EPNL, Joint Nordic method, DIN45681 and Aures' tonality) were compared to identify which of them is able to quantify the detrimental impact of isolated tones and of buzz-saw noise in aircraft sound at takeoff. It was shown that when an isolated tonal component is prominent, all four metrics are able to explain the annoyance ratings. However, when the sound includes a harmonic series of tones with a low fundamental frequency, i.e. buzz-saw noise, only Aures' tonality metric was able to capture it. Therefore, Aures' model for tonality should be considered when trying to capture any kind of audible tones in aircraft sound at takeoff.

It is clear that the metrics which are used to quantify the tonality of sound are based on a spectral analysis and have been developed for stationary sounds. However, aircraft flyovers are non-stationary sounds that contain broadband noise and complex tones [32]. Several

possibilities exist to obtain one overall tonality value for a non-stationary sound. For example, the average or the highest encountered value can be taken. Another possibility is to use tonality exceeded 5% of the time.

2.2.4. ROUGHNESS

A rough sound is a sound with fast loudness fluctuations (50 - 90 Hz). A sound is perceived to be very rough when loudness fluctuations are at around 60 to 70 Hz. A typical example of a rough sound is the sound of a vacuum cleaner. Empirical data related to roughness was reported as early as in 1974 [33]. Various methods exist that try to model the roughness of a sound. One of the earliest methods describing the roughness is given by Zwicker and Fastl [10]. They found that the two characteristics of the ear that influence roughness sensation are: at low frequencies, it is the frequency selectivity of our hearing system whereas at high frequencies, it is the limited temporal resolution. As explained in section 2.2, the masking effect can also be time-varying. This is referred to as temporal masking. Typical examples of sounds producing temporal masking patterns are amplitude modulated (AM) tones or noises, frequency modulated (FM) tones or envelope fluctuations of narrowband noise. The model of roughness proposed by Zwicker and Fastl is based on the temporal masking pattern. The relation between a masker and its temporal masking pattern is schematically illustrated in figure 2.17.

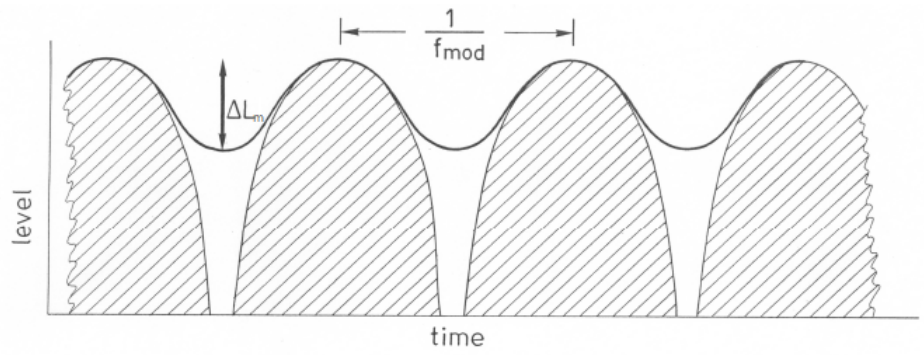


Figure 2.17: Hatched area indicates envelope of amplitude modulated masker. Solid line represents the temporal masking pattern of that masker [10].

The hatched area in figure 2.17 indicates the envelope of a sinusoidally amplitude modulated masker plotted in terms of SPL. The solid line indicates the temporal masking pattern of that sinusoidally amplitude-modulated masker. It can be seen that the interval between two successive peaks of the masker envelope is equal to the reciprocal of the modulation frequency f_{mod}. The masking depth ΔL_m can also be observed in the figure. The masking depth is smaller than the modulation depth due to effects of postmasking, i.e. the decay of psychoacoustic excitation in the hearing system. This observation implies that even if

the SPL of the masker becomes very low, it is still effective in masking other sounds due to effects of postmasking. In this method the roughness R is calculated using equation 2.24, where f_{mod} is the modulation frequency in kHz, ΔL_m is the masking depth in dB/Bark and z is the critical band rate in Bark.

$$R = 0.3 f_{mod} \int_0^{24} \Delta L_m(z) dz \tag{2.24}$$

According to Zwicker and Fastl, the roughness sensation reaches a maximum for modulation frequencies of around 70 Hz. A roughness of 1 asper is achieved by a tone with a center frequency of 1 kHz, SPL of 60 dB and a 100 % amplitude modulation at 70 Hz.

It is noted by Zwicker and Fastl that it might be difficult to quantify the masking depth for sounds typical for practical applications. Then it might be advantageous to derive roughness from the specific loudness pattern, as is done in [34]. Furthermore, in the dissertation of More [12], it is stated that for complex signals, the determination of the modulation frequency can be problematic. Also, the modulation frequency may vary with the critical band rate whereas the equations given above are for a specific modulation frequency. These problems have led to the metric not being standardized yet. Several others method exist which calculate the roughness based on the degree of modulation or in other words modulation depth [35], [36], [37]. The degree of modulation is the ratio of the amplitude of the modulation signal to the amplitude of the carrier signal. The method given by Daniel & Weber [36] is an improvement over Aures' roughness model [35]. Daniel & Weber made some modifications to Aures' roughness model to reduce the deviations between subjective results and calculated results below the just noticeable difference for roughness of 17%. In psychoacoustics, the just noticeable difference is the amount something must be changed in order for a difference to be noticeable. The roughness method of Daniel & Weber was used in the study of Minard & Boussard [13]. In this study, the roughness metric was able to capture changes in buzzsaw noise, which is often heard at take-off.

2.2.5. FLUCTUATION STRENGTH

The fluctuation strength metric is similar to the roughness metric, but deals with slow fluctuations in loudness. The fluctuation strength also shows a bandpass characteristic, meaning that the sensation can only be heard when fluctuations in loudness are from 1 Hz to 20 Hz. Fluctuation strength reaches a maximum when loudness fluctuations are at around 4 Hz. The unit of fluctuation strength is the vacil and a fluctuation strength of 1 vacil is referenced to 1 kHz pure tone with a level of 60 dB 100% amplitude modulated at 4 Hz. Zwicker and Fastl proposed two methods for fluctuation strength: one for tones and another for broadband noise. Again, the temporal variation of the masking pattern, described by the

masking depth, can be used to develop a model for fluctuation strength. It is found that a modulation frequency of 4 Hz plays an important part in fluctuation strength. For higher modulation frequencies, the ear shows the integrative features seen in postmasking, while for lower modulation frequencies the effects of short-term memory become important. For amplitude modulated broadband noise, the temporal masking depth decreases with increasing modulation frequencies. Hence, the temporal masking pattern shows a low-pass characteristic, whereas fluctuation strength has a bandpass characteristic. Taking these effects into account, a relatively simple formula (equation 2.25) is given for the fluctuation strength FS of sinusoidally amplitude modulated broadband noise.

$$FS = \frac{5.8(1.25m - 0.25)[0.05L_{BBN} - 1]}{(f_{mod}/5)^2 + (4/f_{mod}) + 1.5} \tag{2.25}$$

where, m is the modulation depth and L_{BBN} is the level of the broadband noise. Modulated tones show some dependency on frequency. Therefore, for modulated tones, instead of one masking depth, all the masking depths are integrated along the critical band rate scale. Fluctuation strength of amplitude modulated or frequency modulated tones can be approximated with equation 2.26.

$$FS = \frac{0.008 \int_0^{24Bark} \Delta L dz}{(f_{mod}/4) + (4/f_{mod})} \tag{2.26}$$

As was observed by Zwicker and Fastl, fluctuation strength appears to be related to speech and plays a crucial role in the assessment of human speech. At normal speaking rate, usually 4 syllables per second are produced, leading to a fluctuation of the temporal envelope at a frequency of 4 Hz. And as has been stated above, the sensation of fluctuation strength reaches a maximum at modulation frequencies of 4 Hz. As can be expected in nature, the human speech organ produces speech sounds with dominant envelope fluctuations at such a rate for which the human hearing system is most sensitive.

The fluctuation strength is the least important metric when assessing noise annoyance. This might be a reason why there are not many computational models to assess the fluctuation strength of a sound. However, since the roughness and fluctuation metrics are similar to each other, the fluctuation strength metric may be derived from the roughness calculation, as is done by Osses Vecchi et al. [38].

2.2.6. ANNOYANCE ASSESSMENT

To be able to compare the acoustic annoyance based on the SQ metrics with traditional certification metrics, it is important to understand how all the SQ metrics can be combined to determine the acoustic annoyance. A psychoacoustic annoyance (PA) model has been de-

veloped by Zwicker and Fastl [10]. The psychoacoustic annoyance model is given by the following equations:

$$PA = N_5(1 + \sqrt{w_s^2 + w_{FR}^2}) \tag{2.27}$$

$$w_s = 0.25(S - 1.75)log_{10}(N_5 + 10)\ for\ S > 1.75 \tag{2.28a}$$

$$w_s = 0\ for\ S < 1.75 \tag{2.28b}$$

$$w_{FR} = \frac{2.18}{(N_5)^{0.4}}(0.4FS + 0.6R) \tag{2.29}$$

where N_5 is the loudness exceeded 5% of the time, w_s is the term which accounts for sharpness and w_{FR} is the term which accounts for fluctuation strength and roughness.

Their annoyance model incorporates loudness, sharpness, roughness and fluctuation strength. Tonality was not incorporated, since they reasoned that it did not play an important role in the annoyance caused by household products. In aircraft noise however, tonal components are regarded as a very important factor affecting aircraft noise annoyance. For example in the study of Angerer et al. [27], it was shown that after loudness, tonality has the most prominent correlation with the experienced annoyance. Also, in the study of More [12] it was shown that loudness and tonality are the most important metrics that contribute to the experienced annoyance. Therefore, More [12] and also Di *et al.* [39] included tonality in the psychoacoustic model and showed that this modified model yielded the highest correlation with subjective test results. The modified psychoacoustic annoyance model by More is given by equations 2.30 and 2.31. In these equations K_5 is the tonality exceeded 5% of the time and the γ's are coefficients to fit the model results to subjective annoyance ratings.

$$PA_{mod} = N_5(1 + \sqrt{\gamma_0 + \gamma_1 w_s^2 + \gamma_2 w_{FR}^2 + \gamma_3 w_T^2}) \tag{2.30}$$

$$w_{Ton}^2 = [(1 - e^{-\gamma_4 N_5})^2(1 - e^{-\gamma_5 K_5})^2] \tag{2.31}$$

In another study, the combined effect of loudness, tonality and roughness on annoyance ratings of aircraft noise was examined [40]. It was found that the subjects were very sensitive to tonality when both tonality and roughness varied and loudness was kept at the same level. When all three metrics were varied simultaneously, loudness was the major contributor to annoyance. However, tonality and roughness also affected the annoyance ratings.

The attention of the aircraft industry is not only focused on a substantial reduction of the acoustic impact of aircraft, but to a greater extent on a reduction of the fuel consumption. For that reason, new engine architectures such as the ultra high bypass ratio (UHBR) engine and the open rotor engines are being investigated. With the prospect of open rotor engines being mounted on aircraft, tonality will become even more important. Whereas single propellers radiate mostly tonal noise in the propeller plane, two counter-rotating rotors without nacelle radiate many tones over a wide frequency-range due to complex and intense noise interference mechanisms [41]. These open rotor engines also sound rougher than turbofan engines. More [12] found some evidence in his study of an increase in annoyance ratings with increased roughness. This can be attributed to the fact that More also made use of propeller noise measurements, which sound rougher than turbofan engines. Thus with engine architectures such as the open rotor engines, roughness may also become more important. As stated by Minard & Boussard [13] the roughness metric may be used to account for sound sources such as the buzz-saw noise.

Sharpness is the fourth relevant SQ metric for aircraft noise. It is a useful indicator of when high-frequency noise starts to gain prevalence. Aircraft components such as the fan and low-pressure turbine generate high-frequency noise. An increase in sharpness reveals thus that the high-frequency noise generated by the fan has increased relative to the low-frequency aircraft noise components such as airframe noise and jet noise. The final and least relevant metric is the fluctuation strength metric. In the study of More [12] no clear correlation between an increased fluctuation strength and an increase in annoyance was found. Since it is quite similar to the roughness metric, it is still modeled in the AAM.

3

METHODOLOGY

In this chapter the methodology regarding the implemented SQ metrics is stated. In section 3.1 the tonality metric and its implementation in the AAM is given. Section 3.2 elaborates on the roughness metric implemented. In section 3.3 the methodology concerning the fluctuation strength metric is stated.

3.1. AURES' TONALITY

As explained in the previous chapter, many models exist which try to capture the tonal content of a sound. In the automotive industry Aures' Tonality metric has been extensively applied to improve the sound quality of cars [25],[26]. Furthermore, Angerer et al. [27] and More [12] found that the annoyance due to aircraft tonal noise could be predicted with a reasonably high correlation by using Aures' tonality metric. Therefore, it was decided to implement Aures' tonality metric in the AAM.

3.1.1. DETAILS OF TONALITY MODEL GIVEN BY AURES

A schematic overview of the metric is given in figure 3.1. The first step of Aures' method is the extraction of the tonal components. Also, the narrow-band noises need to be extracted, since narrow-band noises with bandwidth less than the critical bandwidth are also perceived as tones by the human ear. The next step is the calculation of the SPL excess, to only consider tones which are aurally relevant. This is done in the same manner as in Terhardt's tonality method. The tonal weighting factor is calculated using three weighting functions, which are explained below.

In figure 3.2 it can be seen that the relative tonality decreases as bandwidth increases. This

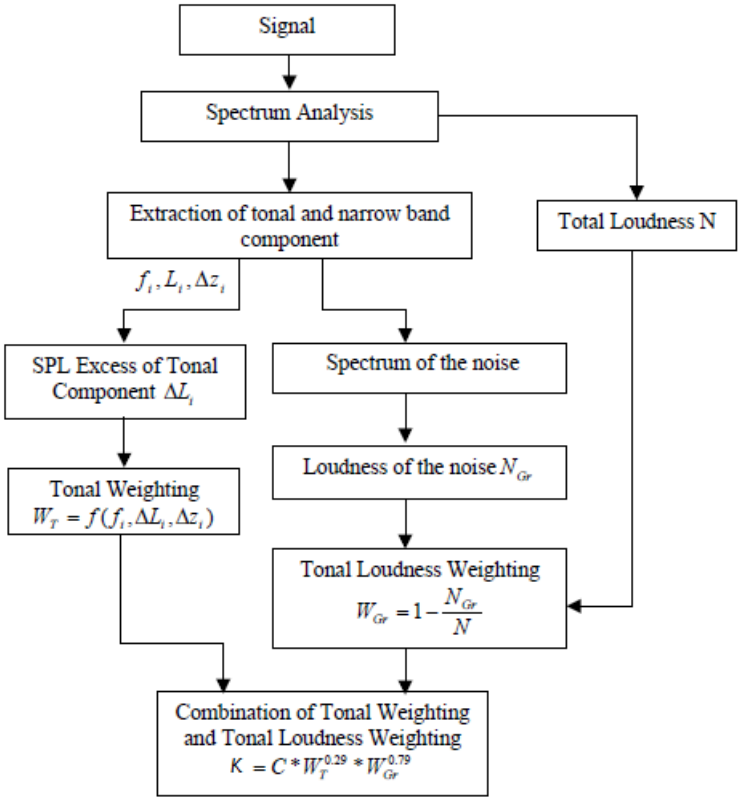


Figure 3.1: Overview of Aures' tonality metric [15]

result was obtained by means of psychoacoustic listening tests.

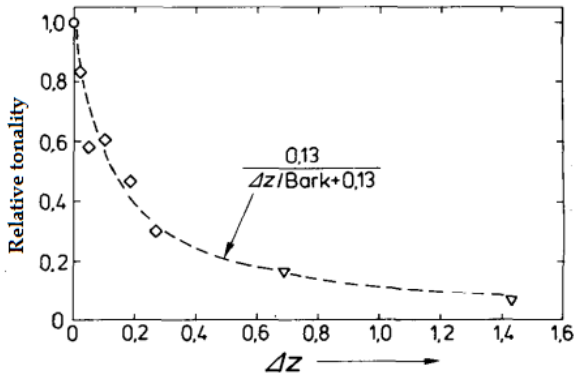


Figure 3.2: Relative tonality versus bandwidth for pure sine tone ∘, narrowband noise with bandwidth of 30 Hz ◊ and bandpass noise with bandwidth of 1 kHz ▽ from Aures' original paper [24]. Dashed line indicates the weighting function w_1.

The results from these psychoacoustic tests were converted in a weighting function by Aures given in equation 3.1, in which Δz is the bandwidth in Bark. This weighting function is also indicated in figure 3.2.

$$w_1(\Delta z) = \frac{0.13}{\Delta z + 0.13} \tag{3.1}$$

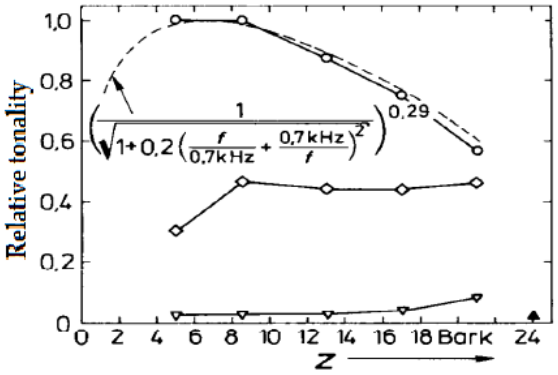


Figure 3.3: Relative tonality versus frequency for pure sine tone ∘, narrowband noise with bandwidth of 30 Hz ◊, bandpass noise with bandwidth of 1 kHz ▽ and highpass noise with cut-off frequency of 2 kHz ▲ from Aures' original paper [24]. The dashed line indicates weighting function w_2.

Usually, the half power points are used to determine the bandwidth Δz. The half power point is that frequency at which the output power drops to half of its peak value. This is at an SPL value 3 dB lower than the peak SPL value.

In figure 3.3 it can be observed that for pure sine tones the tonality has a peak between 5 to 9 Bark. For narrowband noise with a bandwidth of 30 Hz little variation could be seen. And for bandpass noises with a bandwidth of 1 kHz and highpass noise with lower cut-off frequency of 2 kHz almost no tonal perception was observed. The second weighting function used in Aures' tonality is the relation between tonality and frequency of pure tones with no harmonics and can be seen in equation 3.2, where f_i is the center frequency of the i^{th} tonal component in Hz. This equation is also indicated in figure 3.3.

$$w_2(f_i) = \left(\frac{1}{\sqrt{1 + 0.2(\frac{f_i}{700} + \frac{700}{f_i})^2}} \right)^{0.29} \tag{3.2}$$

Aures reasoned that the tonality perception also depends on the level or prominence of the sound. This effect is described by equation 3.3, in which ΔL_i is the excess level of the i^{th} tonal component.

$$w_3(\Delta L_i) = \left(1 - e^{-\frac{\Delta L_i}{15}}\right)^{0.29} \tag{3.3}$$

The three weighting functions are given by equations 3.1-3.3. These three effects form the tonal weighting factor w_T, as given in equation 3.4. It must be noted that first a correction factor ($w_n^{'} = w_n^{1/0.29}$) is applied to the three weighting factors, which was introduced by Aures to match the subjective responses more closely.

$$w_T = \sqrt{\sum_{i=1}^{n} [w_1'(\Delta z_i)\, w_2'(f_i)\, w_3'(\Delta L_i)]^2} \tag{3.4}$$

Aures also found by means of listening tests that the tonal weighting was not accurate enough for sounds composed of both broadband and tonal noise. Therefore, parallel to the tonal weighting factor, a loudness weighting factor is calculated which is based on the signal-to-noise ratio. First, the loudness of the noise without tones N_{Gr} is determined. This can be used to calculate the loudness weighting factor with equation 3.5. In this equation N_{total} represents the total loudness level.

$$w_{Gr} = 1 - \frac{N_{Gr}}{N_{total}} \tag{3.5}$$

Finally, both the tonal and loudness weightings are combined into equation 3.6. In this equation, c is a calibration term to give a 1 kHz tone of 60 dB SPL a tonality of one. The unit of tonality is tonality unit (t.u.).

$$K = c \cdot w_T^{0.29} \cdot w_{Gr}^{0.79} \tag{3.6}$$

3.1.2. IMPLEMENTATION IN THE AAM

In this section, the implementation of Aures' tonality in the AAM is given in detail. The code is largely based on the code given in the dissertation of Hastings [29]. After a file is loaded in the AAM, the signal is divided into time-blocks of 0.2 seconds. Hereafter, for each time-block the SPL is calculated. A flow diagram for the tonality calculation is given in figure 3.4.

After the initialization, the tonal components are extracted and bandwidth & center frequency of each tonal component is calculated. The bandwidth of a tonal component is estimated using the half-power points. First the peak level of the tonal component is determined and then a inspection is conducted to the left and right of the peak to determine the closest spectral points where the SPL has dropped by at least 3 dB from the peak level. The center frequency of the tonal component is not equal to the frequency corresponding to the peak level, but is calculated using equation 3.7.

$$f_c = \sqrt{f_l \cdot f_u} \tag{3.7}$$

where f_l is the lower cut-off frequency and f_u is the upper cut-off frequency of the tonal component. This is followed by the generation of the tonal spectrum and noise spectrum. The tonal spectrum contains only the tones found in previous steps, while the noise spectrum is the original spectrum with the tones filtered out. The generation of the noise spectrum requires the calculation of the noise floor. The noise floor is defined as that level for

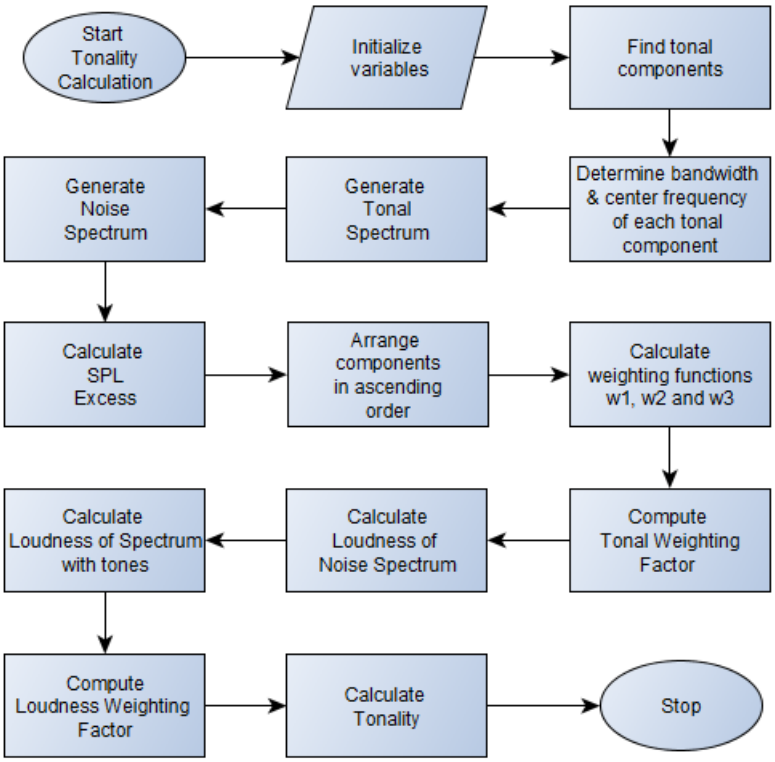


Figure 3.4: Flow diagram of tonality calculation

which 90% of the spectrum in the critical band centered around the tonal component will be above. The noise floor below the tonal component is determined by using a linear slope to connect the tonal component endpoints. The next step in the tonality calculation is the determination of the SPL excess of each tonal components. This is done according to equation 2.12. Only the components with a SPL excess > 0 dB are relevant. Then the center frequency, bandwidth and excess level of each tonal components is arranged in ascending order. The weighting functions $w1, w2$ and $w3$ are computed according to equations 3.1 - 3.3 and are used to compute the tonal weighting factor. This is followed by the loudness calculation, which is done according to DIN45631 [42], for both the tonal spectrum and noise spectrum. With the loudness of both spectra, the loudness weighting factor is calculated. Finally, the tonality can be computed. This tonality calculation is repeated for each time block. For more details about the tonality implementation in Matlab, see [29].

3.2. ROUGHNESS

As stated before, the roughness method of Daniel & Weber (hereafter referred to as D&W's model) is able to capture changes in buzzsaw noise. For that reason, D&W's model will be implemented in the AAM. In section 3.2.1 details of the roughness model proposed by D&W are given. In section 3.2.2 the implementation of the model in the AAM is discussed.

3.2.1. DETAILS OF ROUGHNESS MODEL GIVEN BY DANIEL & WEBER

The structure of the model proposed by Daniel & Weber [36] is illustrated in figure 3.5. This model is based on the roughness model of Aures [35]. Some modifications are made to Aures' model to reduce the deviations between subjective results and calculated results below the just noticeable difference for roughness of 17%.

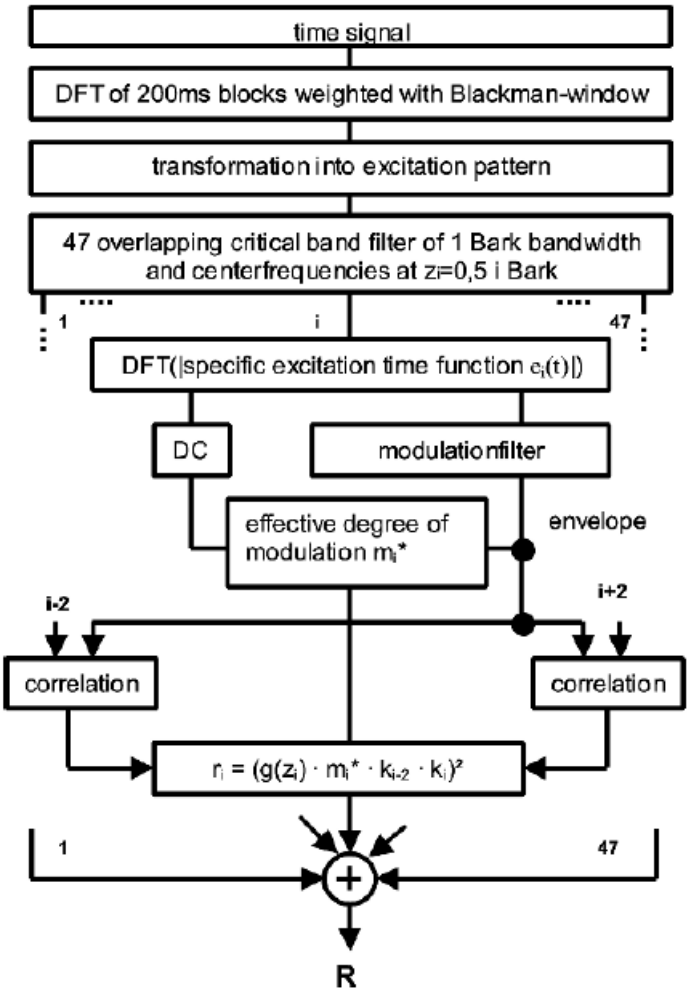


Figure 3.5: Overview of roughness method proposed by Daniel and Weber [36]

The Discrete Fourier Transform (DFT) of the time signal is taken using analysis windows of 200ms weighted by a Blackman window. Then, the spectrum is multiplied by a factor a_0

representing the transmission between free field and the peripheral hearing system. This transmission factor is indicated by the solid line in figure 3.6. The dotted line in the figure represents the transmission factor for diffuse-field condition a_{0D}.

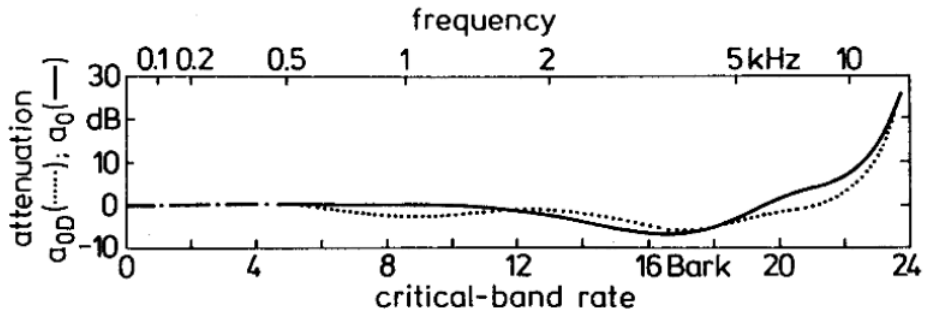


Figure 3.6: Transmission factor for free-field condition (a_0, solid curve) and diffuse-field condition (a_{0D}, dotted curve) [10]

This is followed by a transformation into an excitation pattern with the slopes of the excitation pattern defined by equation 2.15. Subsequently, specific excitations are calculated within 47 overlapping critical bands (channels) with a bandwidth of 1 Bark. Using an inverse DFT yields 200ms long specific excitation time functions $e_i(t)$. Then a weighting function $H_i(f_{mod})$ is applied to model the bandpass characteristic of roughness. Since 47 overlapping critical bands are used, 47 weighting functions are applied. The weighting functions $H_2(f_{mod})(z_2 = 1\ Bark)$, $H_{16}(f_{mod})(z_{16} = 8\ Bark)$ and $H_{42}(f_{mod})(z_{42} = 21\ Bark)$ are drawn in figure 3.7.

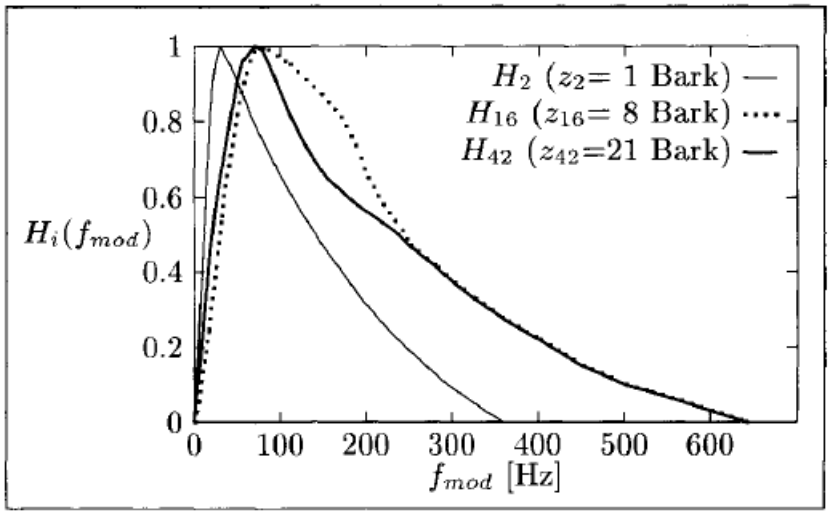


Figure 3.7: Weighting functions H_2, H_{16} and H_{42} for the roughness metric [36]

The weighting functions drawn in figure 3.7 are basic in the sense that the weighting functions in the other channels can be derived from them. This can be done in the following way:

for i = 1,2,...,4:	$H_1 = H_2 = H_3 = H_4$
for i = 6,8,...,16	$H_{i-1} = H_i$ and
	H_i is linearly interpolated between H_4 and H_{16}
for i = 17,18,19,20	$H_i = H_{16}$
for i = 22,24,...,42	$H_{i-1} = H_i$ and
	H_i is linearly interpolated between H_{20} and H_{42}
for i = 43,44,...,47	$H_i = H_{42}$ and

The weighted excitation envelope can then be calculated using equation 3.8, in which $e_i(t)$ is the specific excitation time function.

$$h_{BP,i}(t) = IFFT\{H_i(f_{mod}) \cdot FFT(|e_i(t)|)\} \tag{3.8}$$

The DC-value is calculated according to equation 3.9.

$$h_{0,i} = |\tilde{e}_i(t)| \tag{3.9}$$

After this the RMS value of the weighted excitation envelope is divided by the DC value of each original filtered signal to give a generalized modulation depth m^*. The modulation depth is limited to 1 to avoid unrealistic high roughness values that might be produced by pulses in the excitation envelope:

$$m_i^* = \begin{cases} \frac{\tilde{h}_{BP,i}(t)}{h_{0,i}}, & if\ \frac{\tilde{h}_{BP,i}(t)}{h_{0,i}} \leq 1 \\ 1, & if\ \frac{\tilde{h}_{BP,i}(t)}{h_{0,i}} > 1 \end{cases}$$

The specific roughness can then be calculated using equation 3.10.

$$r_i = (g_R(z_i) \cdot m_i^* \cdot k_{i-2} \cdot k_i)^2 \tag{3.10}$$

where g_R represents calibration factors introduced to account for the dependency of roughness on carrier frequency, k_{i-2} is the crosscorrelation coefficient between the envelopes of channel $i-2$ and i and k_i is the crosscorrelation coefficient between the envelopes of channel i and $i+2$. The calibration factor g_R versus center frequency z is plotted in figure 3.8. The total roughness is then equal to the sum of specific roughnesses.

$$R = 0.25 \sum_{i=1}^{N} r_i \tag{3.11}$$

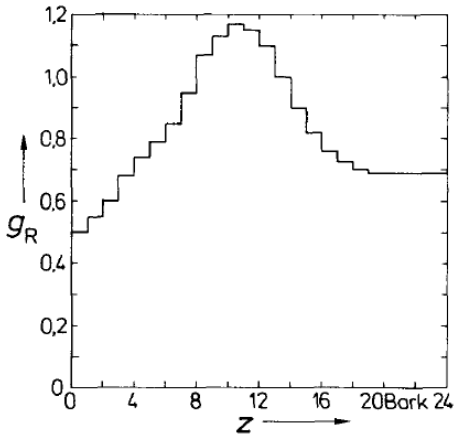


Figure 3.8: Calibration factors for the roughness metric [36]

3.2.2. IMPLEMENTATION IN THE AAM

In this section the implementation of D&W's roughness in the AAM is given. The implementation of this metric was done by Schrader in his dissertation [43]. The code was adapted and corrected to be able to integrate it in the AAM. The corrections concern changes in weighting functions and calibration factors and will be explained in this section. As in the tonality calculation, time-blocks of 0.2 seconds are used. A flow diagram of the roughness calculation is depicted in figure 3.9.

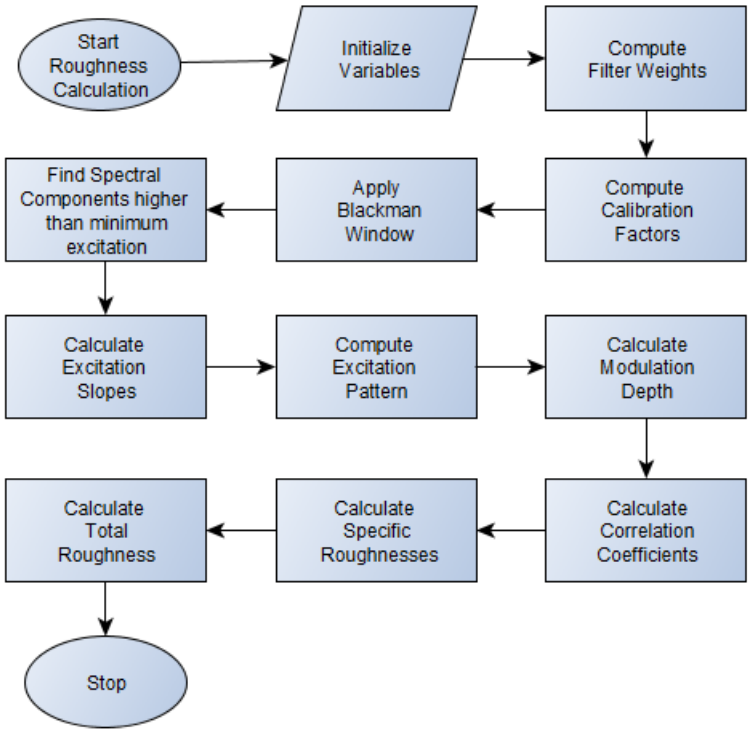


Figure 3.9: Flow diagram of roughness calculation

After the initialisation of variables, the filter weights which model the bandpass dependency on modulation frequency are calculated. The weighting functions in [43] were defined in the following way:

$$H_i = \begin{cases} H_2, for\ 1 \leq i \leq 4 \\ H_6, for\ 5 \leq i \leq 15 \\ H_{16}, for\ 17 \leq i \leq 20 \\ H_{21}, for\ 22 \leq i \leq 41 \\ H_{42}, for\ 43 \leq i \leq 47 \end{cases}$$

It can be seen that this is not in accordance to the weighting functions given in the paper of Daniel & Weber [36]. Therefore, new weighting functions were used that did agree with the weighting functions given in [36]. The next step in the calculation is the computation of the calibration factors g_R. Initially, the calibration factors were similar to the ones given in Aures' model [35]. However, they were modified to tune the results as close as possible to subjective results of roughness of AM-tones. In figure 3.10 the original calibration factors and those used in the implementation are plotted.

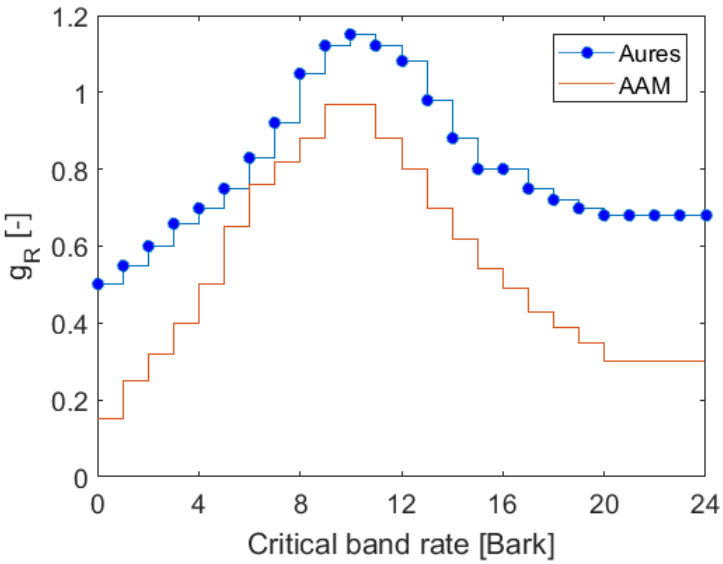


Figure 3.10: Implemented calibration factors and calibration factors given by Aures for roughness metric

In the roughness calculation, a Blackman window is applied to the signal, the transmission factor a_0 is applied and the SPL is calculated. This spectrum is then used to find the spectral components that are higher than the minimum excitation. These components are used to calculate the excitation slopes and to determine the excitation pattern. Subsequently, the modulation depths m^* and correlation coefficients k are calculated. The specific roughnesses are then calculated in the following way:

$$r_i = \begin{cases} (g_R(z_i) \cdot m_i^* \cdot k_i)^2, for\ i = 1,2 \\ (g_R(z_i) \cdot m_i^* \cdot k_{i-2} \cdot k_i)^2, for\ i = 3,4,...,45 \\ (g_R(z_i) \cdot m_i^* \cdot k_{i-2})^2, for\ i = 46,47 \end{cases}$$

The final step is the calculation of total roughness according to equation 3.11.

3.3. FLUCTUATION STRENGTH

The fluctuation strength is the least important metric when assessing noise annoyance [12] [7]. This might be a reason why there are not many computational models to assess the fluctuation strength of a sound. Taking advantage of the physical similarity between roughness and fluctuation strength, a model was developed by Osses Vechi et al. of Eindhoven University of Technology [38]. This model is referred to as TU Eindhoven's model henceforth and is explained in detail in the following section.

3.3.1. DETAILS OF FLUCTUATION STRENGTH MODEL

TU Eindhoven's model was adapted from the roughness model given by Daniel & Weber [36]. A schematic overview of this method can be seen in figure 3.11.

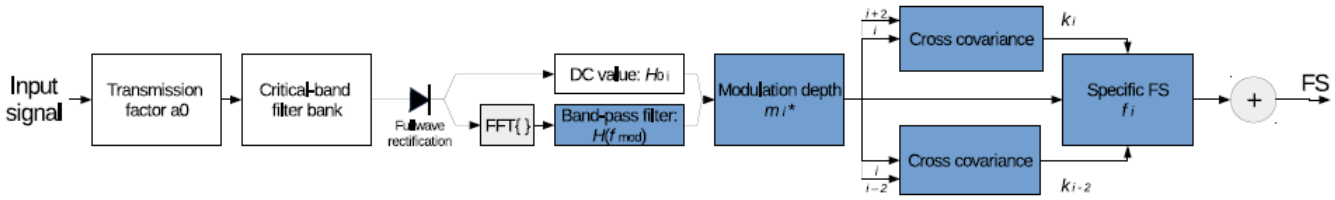


Figure 3.11: Overview of fluctuation strength method proposed by Osses Vecchi et al. [38]

It can be seen that this model is based on the models of roughness given in [36], [35]. The total fluctuation strength (FS) is equal to the sum of the specific fluctuation strengths, as depicted in equation 3.12.

$$FS = \sum_{i=1}^{N} f' = C_{FS} \cdot \sum_{i=1}^{N} (m_i^*)^{p_m} \cdot |k_{i-2} \cdot k_i|^{p_k} \cdot (g_{FS}(z_i))^{p_g} \tag{3.12}$$

The parameters C_{FS}, p_m, p_k and p_g are constants and are optimized to fit their experimental data. The experimental data contained artificial stimuli (AM and FM tones and AM broadband noise) and everyday sounds such as male/female speech and musical instrument sounds. The following values were found for the parameters: $C_{FS} = 0.249$, $p_m = p_k = 1.7$ and $p_g = 1$.

3.3.2. IMPLEMENTATION IN THE AAM

The implementation of the fluctuation strength metric in the AAM is based on the implementation of the roughness metric. For the fluctuation strength metric, the signal is divided into time-blocks of 2 seconds, in contrast to the time-blocks of 0.2 seconds for roughness and tonality. This is because a higher frequency resolution is required for fluctuation strength and the frequency resolution is inversely proportional to the snapshot. No flow diagram is given in this section, since the calculation steps are similar to the roughness calculation.

The weighting functions are derived from the weighting functions in the roughness calculation, such that the highest values for the weighting functions occur at a modulation frequency of 4 Hz. The calibration factors g_{FS} are kept same as the ones in the roughness calculation (g_R). In the roughness model the modulation depth was limited to 1 to avoid unrealistic high roughness values. In [38] it is argued that this is not needed for fluctuation strength, and a compression ratio of 3:1 is suggested for modulation depths higher than 0.7 rather than a limitation. For example, an input modulation depth of 0.85, which is 0.15 units higher than the threshold of 0.7, will be compressed to a output modulation depth of 0.75.

4

VALIDATION

This chapter deals with the validation of the implemented metrics. In section 4.1 the validation of the tonality metric is given. This is followed by the validation of the roughness metric in section 4.2. Finally, the validation of the fluctuation strength metric is given in section 4.3.

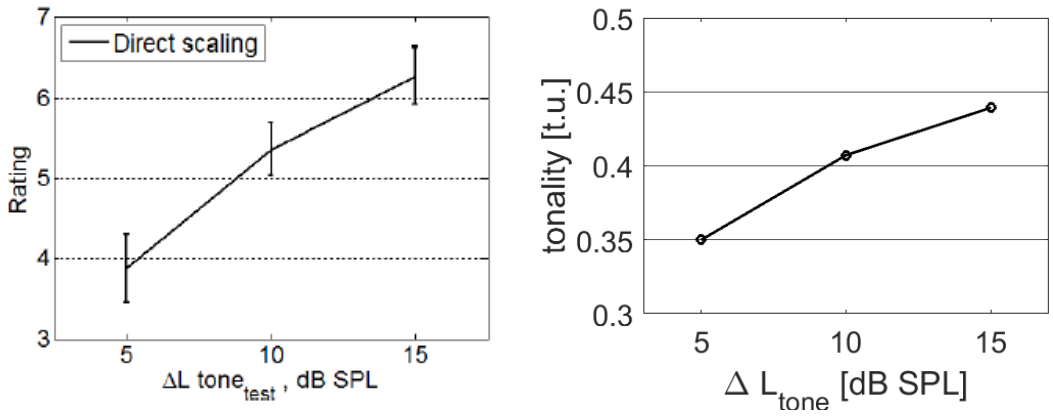
4.1. VALIDATION OF TONALITY METRIC

As has been stated before one t.u. is referenced to a pure sine tone with a SPL of 60 dB and a frequency of 1000 Hz. First of all, it is inspected whether this is true for the implemented metric. A stimuli was created with a SPL of 60 dB and a tone frequency of 1000 Hz. Loading this stimuli in the AAM gave a t.u. value of 1.007 tonality unit. A bandwidth correction has been applied in the tonality metric to ensure that a pure sine wave has a bandwidth of zero.

The validation has been done using publicly available data. First, three effects were studied on stimuli consisting of a tone and background noise: the influence of tone level (ΔL_{tone}) on tonality perception, the influence of tone frequency (f_{tone}) on tonality perception and the influence of noise bandwidth on tonality perception. From literature, it is known that a higher tone emergence level and higher tone frequency increase the perceived tonality. This was also the result in a study performed by Dendievel et al. [44]. In the study, 18 stimuli, with varying bandwidth of background noise, were evaluated by 9 subjects using three listening test methods. The three listening methods used were the following: direct scaling, two alternative forced choice and an iso-tonality equalization method. Results generated using the AAM will be compared to the results of the direct scaling method used in the study. The results from the other listening test methods were similar to the results obtained

using the direct scaling method. Background noise was created by adding white uniform noise. The desired spectral width is obtained by means of a Butterworth filter.

First, the influence of tone emergence level on tonality perception was investigated. The tone emergence level is defined as the difference between the level of the tone and that of the noise centered on the tone frequency. The tonality was computed using three different tone emergence levels: 5, 10 and 15 dB SPL. A tone frequency of 1200 Hz and a noise bandwidth of 2 Bark was used. In figure 4.1 the results can be seen.



(a) Ratings of listening tests

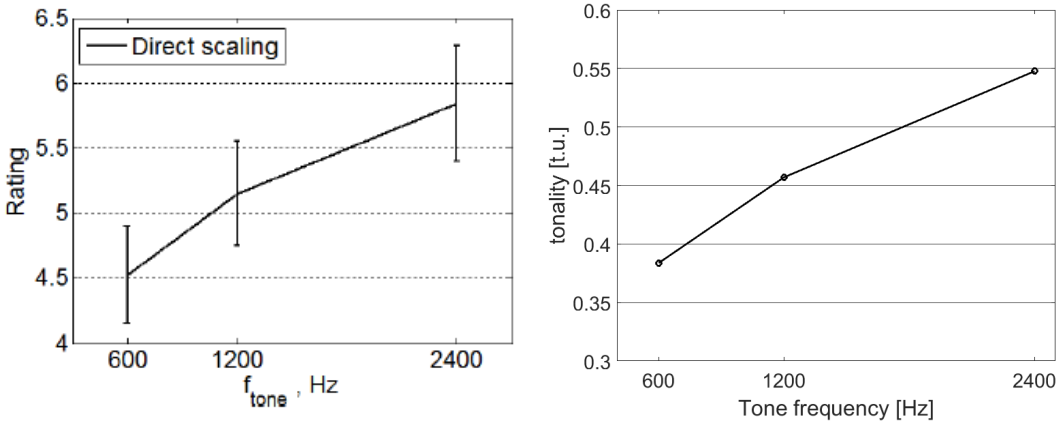
(b) Aures' tonality (AAM)

Figure 4.1: Influence of tone emergence level on tonality

In figure 4.1a, the result of the direct scaling method used in the study of Dendievel et al. can be seen. The perceived tonality increases when tone emergence level increases. In figure 4.1b, the results of Aures' tonality implemented in the AAM are given. Again, the tonality increases when increasing the tone emergence level. Thus this is in line with the results of the listening tests.

Next, the influence of tone frequency on tonality perception was examined. The tonality was computed for three frequencies: 600 Hz, 1200 Hz and 2400 Hz. The noise bandwidth is fixed at 2 Bark. Since the critical bands get wider for higher frequencies, the noise bandwidth in Hz is greater for higher frequencies. The tone emergence level is fixed at 20 dB. In figure 4.2 the results can be seen.

In figure 4.2a it is observed that the ratings increase with increasing tone frequency. This increase can also be observed in figure 4.2b, where the tonality is computed using the AAM. These results are thus also in line with each other.



(a) Ratings of listening tests

(b) Aures' tonality (AAM)

Figure 4.2: Influence of tone frequency on tonality

The influence of background noise on tonality perception is the last effect to be investigated in the study of Dendievel et al. [44]. The authors of the paper argue that most tonality indicators are not capable to correctly take this effect into account. From the listening tests it was established that the tonality perception decreases when the background noise bandwidth increases (see figure 4.3).

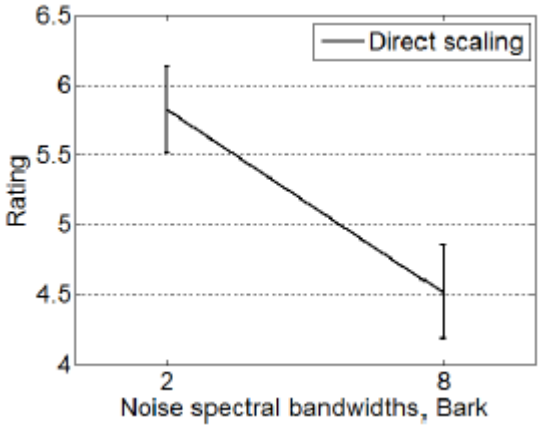


Figure 4.3: Ratings of listening tests related to the effect of noise bandwidth

As stated above, one of the aims of the study was to investigate whether tonality indicators take the influence of background noise bandwidth into account. Five tonality indicators were considered for this investigation: maximum prominence ratio (PR), tone-to-noise ratio (TNR), the DIN45681 standard, Tonal Audibility of ISO1996-2 and Aures' tonality (average over time). The tone emergence level is fixed at 10 dB and stimuli have a duration of 60 seconds. Background noise with a bandwidth of 2 Bark, 4 Bark and 8 Bark is used. In figure 4.4 the results of the study are displayed for a tone frequency of 1200 Hz.

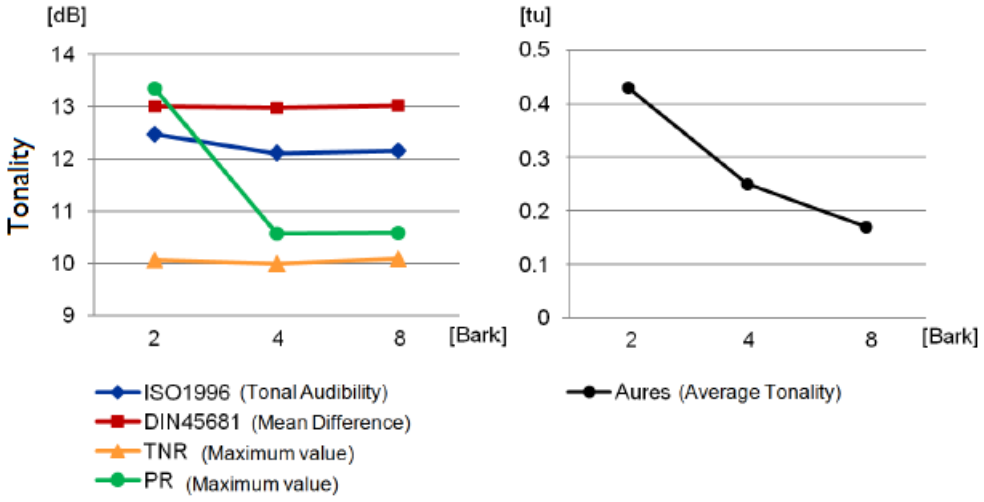


Figure 4.4: Influence of background noise bandwidth on five tonality indicators [44].

It is evident that only Aures' tonality metric is able to take the influence of background noise bandwidth into account. This is due to the fact that a weighting is applied that is based on the relative loudness of the tone and noise (W_{gr}). The other four tonality indicators stay virtually the same with varying noise bandwidth. Only the value for PR for a bandwidth of 2 Bark is higher since the indicator calculates the energy ratio between the critical band centered at the tone and two adjacent critical bands.

The investigation was also conducted for the tonality indicator implemented in the AAM and compared to the results of the study (figure 4.5).

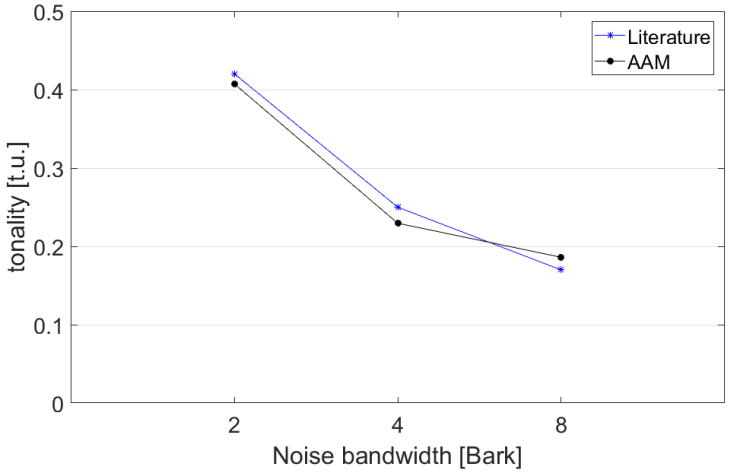
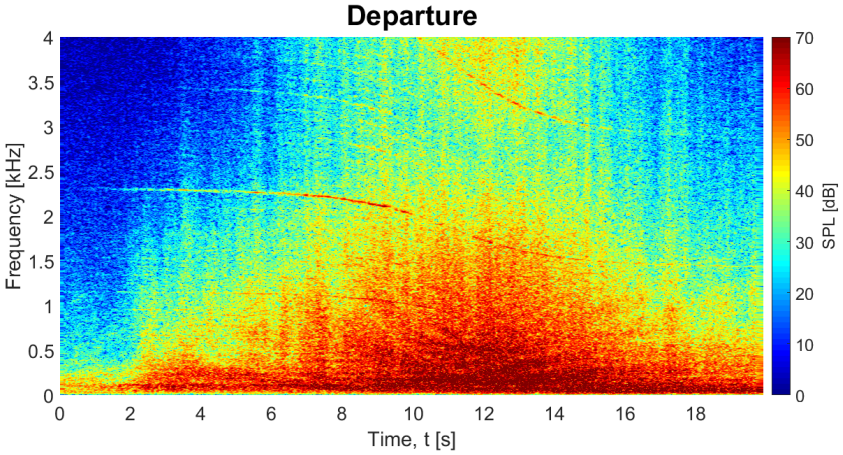


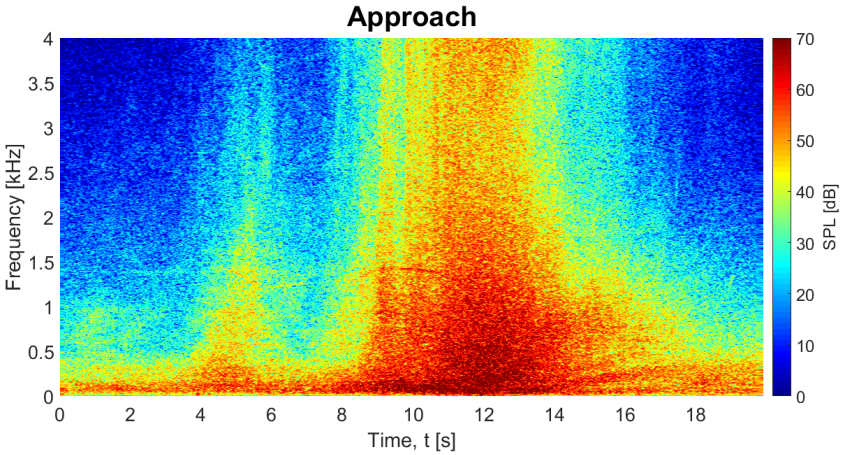
Figure 4.5: Comparison of noise bandwidth effect between Aures' tonality implemented in AAM and and result for Aures' tonality from [44]

It can be seen that both the results from the paper and from the AAM show the same trend: tonality decreases when increasing the background noise bandwidth. A small difference can be observed between the two curves. This may be attributed to the fact that the added uniform noise was slightly different, or that the cutt-off frequencies for the Butterworth filter were slightly different. Nevertheless, it has been established that the tonality metric implemented in the AAM is able to take the influence of background noise bandwidth into account.

The last part of the validation of the tonality will be done by making use of aircraft flyover measurements. A comparison will be made for the tonality values between departure and approach. It is expected that the tonality value for departure is higher due to the buzz-saw tones present during take-off. The comparison is made for the Boeing 737-700 and the Embraer 190. The spectrograms for the Boeing 737-700 for departure and approach can be seen in figure 4.6.



(a) Departure spectrogram of Boeing 737-700



(b) Approach spectrogram of Boeing 737-700

Figure 4.6: Departure and approach spectrograms of Boeing 737-700

By looking at the spectrograms, it becomes clear that the aircraft is more tonal during departure. Multiple strong tones and harmonics can be observed in the departure spectrogram. This is mainly due to the presence of buzz-saw tones during take-off. This fact is reflected by the K_5-values for both aircraft in table 4.1.

Aircraft	**Departure**	**Approach**
B737-700	0.2945	0.2352
ERJ-190	0.3167	0.2451

Table 4.1: Comparison of K_5-values in t.u. for departure and approach

In this section the validation of tonality has been done using data from literature and by comparing approach and take-off aircraft flyovers. The tonality metric implemented in the AAM is able to correctly capture the effect of tone emergence level, tone frequency and background noise. Applied to aircraft flyovers, the results complied with the expectation of higher tonality at take-off.

4.2. VALIDATION OF ROUGHNESS METRIC

Once again it is first inspected whether the reference value of one asper is calculated correctly. One asper is referenced to a 60 dB SPL 1000 Hz tone which is 100% modulated at 70 Hz. Loading this stimuli in the AAM gave a roughness value of 1.0009 asper.

The roughness model of Daniel & Weber was optimized to reproduce the subjectively found characteristic bandpass dependence on the modulation frequency of amplitude-modulated tones. This was also done for the roughness code implemented in Matlab by adjusting the g-factors in the model. Therefore, it is attempted to reproduce figure 3 in [36] which illustrates the roughness as a function of modulation frequency for amplitude-modulated tones. The results can be seen in figure 4.7. More information regarding amplitude modulation can be found in appendix A.

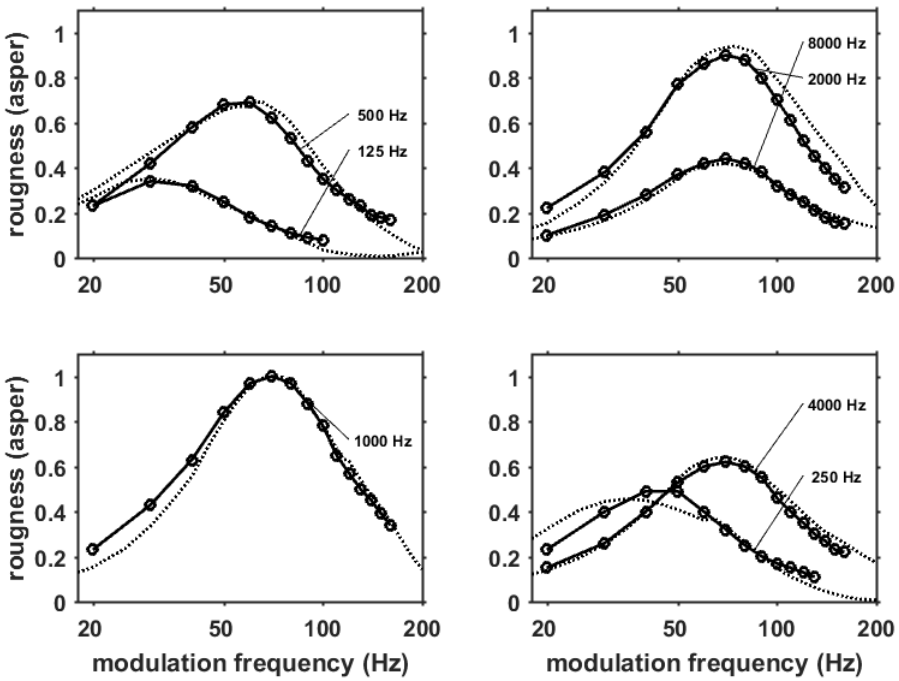


Figure 4.7: Roughness of amplitude modulated tones. Calculated roughness (···) and roughness data obtained from listening tests (-o-) of amplitude-modulated tones as a function of modulation frequency f_{mod} and carrier frequency f_c with modulation depth = 1 and SPL = 60 dB.

The results are given for the following carrier frequencies: 125 Hz, 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz and 8000 Hz. In the plot the results calculated with the AAM and the measured data published by Zwicker and Fastl [10] are given. It can be seen that there is good agreement between the modelled and the subjective data from listening tests. In the lower left plot it can be seen again that a value of 1 asper is found for a modulation fre-

quency of 70 Hz. The largest discrepancy can be seen for $f_c = 250Hz$ and $f_{mod} < 50Hz$ and for $f_c = 2000Hz$ and $f_{mod} > 70Hz$. However, it has been demonstrated that the bandpass dependence on modulation frequency has been modelled correctly in the roughness metric.

Additionally, the influence of modulation depth on roughness is investigated. In figure 4.8 the calculated roughness is plotted as a function of modulation depth together with the power law $R = 1.36m^{1.6}$. The power relation is given by Zwicker and Fastl [10] and is an approximation to the subjective data measured.

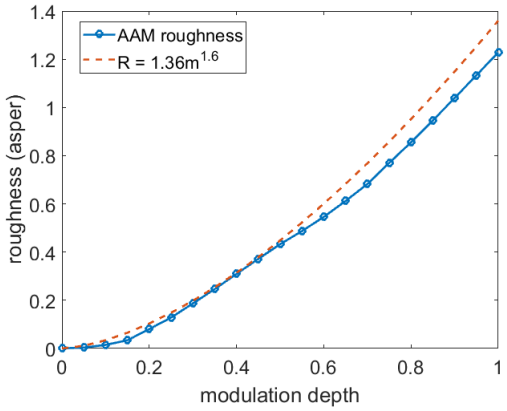
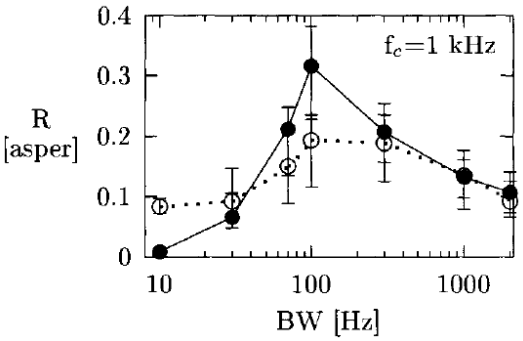


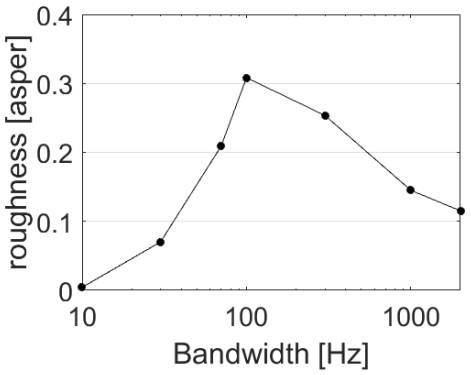
Figure 4.8: Roughness versus modulation depth

There is a fairly good agreement between the model results and the approximation. For modulation depths greater than 0.5, an increasing discrepancy can be observed. Nonetheless, the differences are smaller than the just noticeable difference of 17%.

In the paper of Daniel & Weber [36] roughness is calculated for varying noise bandwidth. The result for a center frequency of 1000 Hz and SPL of 70 dB is given in figure 4.9a. In the figure roughness had been calculated for the following bandwidths: 10 Hz, 30 Hz, 70 Hz, 100 Hz, 300 Hz, 1000 Hz and 2000 Hz. Measured data derived from [35] is also plotted in the figure (circles connected by dots). The bandpass noises used in the calculations are simulated by digital bandpass filtering of normal distributed random numbers. For each bandwidth, 21 roughness calculations are carried out and the median is plotted in the figure together with the quartiles. The roughness as a function of bandwidth was also calculated using the AAM and the result is plotted in 4.9b.



(a) Calculated roughness (-•-) and roughness obtained using listening tests (·∘·) as a function of bandwidth given in paper D&W [36]



(b) Calculated results with the AAM

Figure 4.9: Roughness versus bandwidth

It can be seen that the calculated results from figure 4.9a and 4.9b are very similar. There is a fairly good agreement between the subjective data obtained using listening tests given in figure 4.9a and the calculated data in figure 4.9b. Thus it can be concluded that the model is able to correctly predict roughness for unmodulated bandpass noise.

Although Daniel & Weber's roughness model has been optimized to fit subjective measurements of AM-tones, the model also shows good agreement between measured and calculated data for FM-tones, as can be seen in figure 4.10. More information regarding frequency modulation can be found in appendix A.

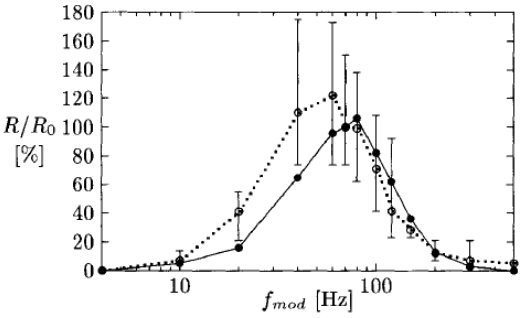


Figure 4.10: Calculated relative roughness (-•-) and relative roughness data obtained using listening tests (·∘·) versus modulation frequency for FM-tones given in paper D&W [36]

The relative roughness $\frac{R}{R_0}$ (relative to a FM tone with f_{mod} = 70 Hz) is plotted as a function of modulation frequency for a 1600 Hz FM tone with a frequency deviation Δf of 800 Hz and SPL of 60 dB. Differences between calculated and subjective data can be observed for low modulation frequencies. For high modulation frequencies the data correspond very well with each other. This figure was reproduced using the AAM and the result can be seen

in figure 4.11.

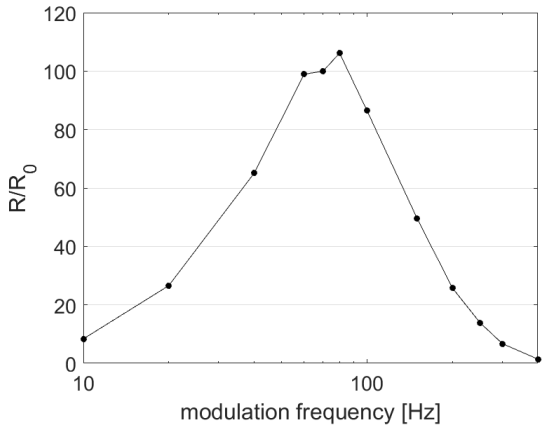


Figure 4.11: Relative roughness versus modulation frequency calculated with the AAM

This plot corresponds to the calculated result as given in figure 4.10. The bandpass characteristic of roughness with the peak at around 70-80 Hz can be detected in the figure.

The roughness metric will also be validated using commercial software: Artemis Suite of HEAD acoustics[8]. In this software the roughness metric of Sottek [45] is implemented. Sottek's roughness metric is similar to D&W's roughness metric in the way that it calculates total roughness by summing up specific roughnesses in each critical band. It is thus expected that the roughness values of Sottek's method and D&W's method are comparable.

In table 4.2 the roughness values are calculated for 20 aircraft flyover measurements.

Table 4.2: R_5 for 20 aircraft flyover measurements

Measurement	AAM (R_5)	Artemis (R_5)	**Measurement**	AAM (R_5)	Artemis (R_5)
1	0.0815	0.0706	11	0.0948	0.0718
2	0.0924	0.0766	12	0.0834	0.0693
3	0.0755	0.0753	13	0.0835	0.0755
4	0.0962	0.0755	14	0.0959	0.0789
5	0.0696	0.0715	15	0.1008	0.0815
6	0.1137	0.0782	16	0.1054	0.0827
7	0.1106	0.0810	17	0.0842	0.0795
8	0.0801	0.0697	18	0.9423	1.1
9	0.0901	0.0667	19	0.1064	0.0768
10	0,0838	0,0715	20	0,0784	0,0705

[8]https://www.head-acoustics.com/eng/nvh_artemis_suite.htm

It can be observed that the roughness values for AAM and Artemis have the same order of magnitude. However, there is more variance in the values calculated with the AAM. The absolute values are not identical for AAM and Artemis when comparing the results for a specific flyover measurement. This is as expected, since the methods used are different from each other. Therefore, the ratio between different results should be considered more important.

It can be seen that all values are in the range of 0.06 - 0.12 asper. However, one outlier (measurement 18) can be seen. The R_5 values calculated for this measurement are 0.94 asper (AAM) and 1.1 asper (Artemis). This indicates that the roughness metric implemented in the AAM is also able to correctly detect strong fluctuations in sound. Roughness over time is plotted in figures 4.12 and 4.13 for Artemis and AAM roughness respectively. The plots are very similar. High values for roughness can be observed for the first two seconds, which is caused by an unknown buzzing disturbance in the measured flyover. Furthermore, in both plots a peak of approximately 0.4 asper can be observed at $t = 5$ seconds. Hereafter, the plots are relatively flat and an increase can be seen towards the end.

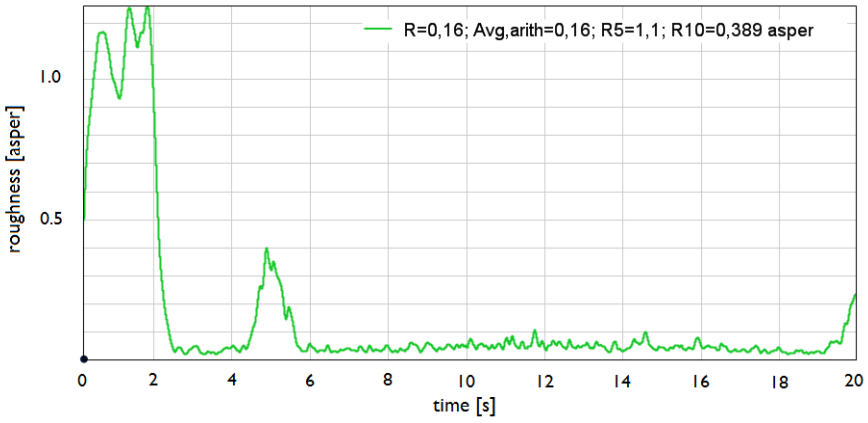


Figure 4.12: Roughness vs time calculated with Artemis for measurement 18

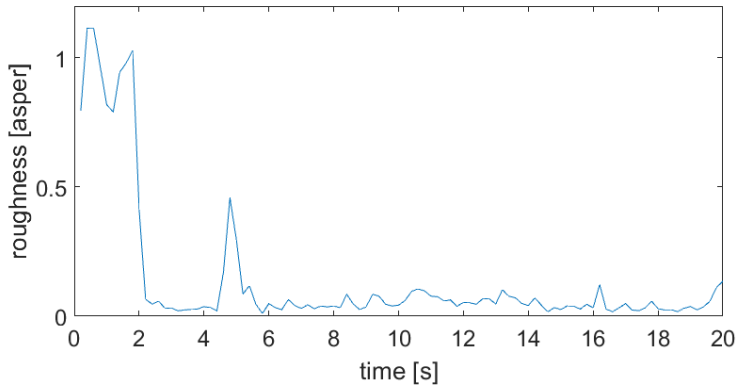


Figure 4.13: Roughness vs time calculated with AAM for measurement 18

The validation using Artemis is also performed by making use of "everyday" sounds. The sounds include for example a vacuum cleaner, a chainsaw and police siren. Roughness exceeded 5% of the time is calculated with both the AAM and Artemis and a linear regression is drawn through these points, as can be seen in figure 4.14. High correlation (R^2 = 0.8024) between the results can be observed.

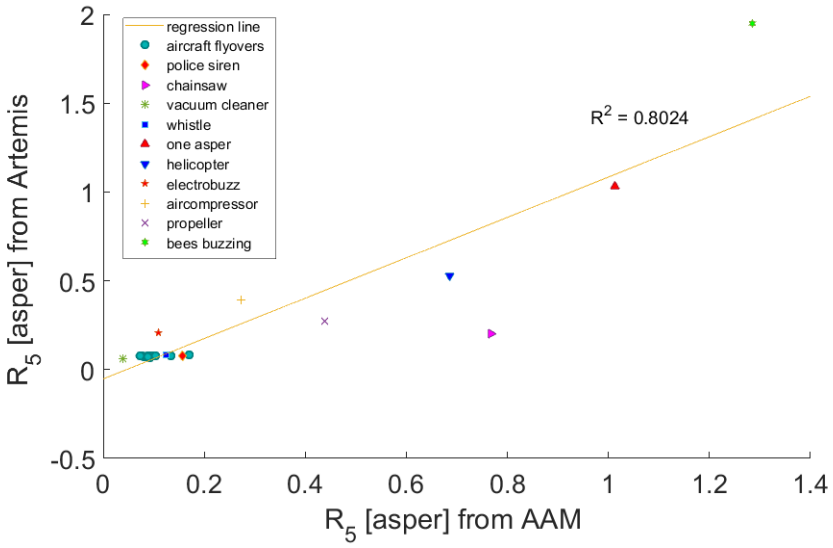


Figure 4.14: Correlation between results generated with the AAM and Artemis

The validation of roughness has been done using data found in literature and using Artemis. There is a good agreement between the results presented in literature and the results generated with the roughness model implemented in the AAM. Although the roughness methods in Artemis and the AAM differ from each other, the results generated using Artemis showed good agreement with the results generated with the AAM.

4.3. VALIDATION OF FLUCTUATION STRENGTH METRIC

First of all, it is examined whether the reference value of one acum is computed correctly. One acum is referenced to a 60 dB SPL 1000 Hz tone which is 100% modulated at 4 Hz. Loading this stimuli in the AAM gave a fluctuation strength value of 1.0038.

The bandpass dependency on modulation frequency of the fluctuation strength metric will now be tested using amplitude-modulated (AM) tones, frequency-modulated (FM) tones and AM broadband noise. The results of the AAM will be compared with results of the model of TU Eindhoven [38] and results found in literature which were obtained using perceptual experiments [10]. The results for AM tones are given in figure 4.15.

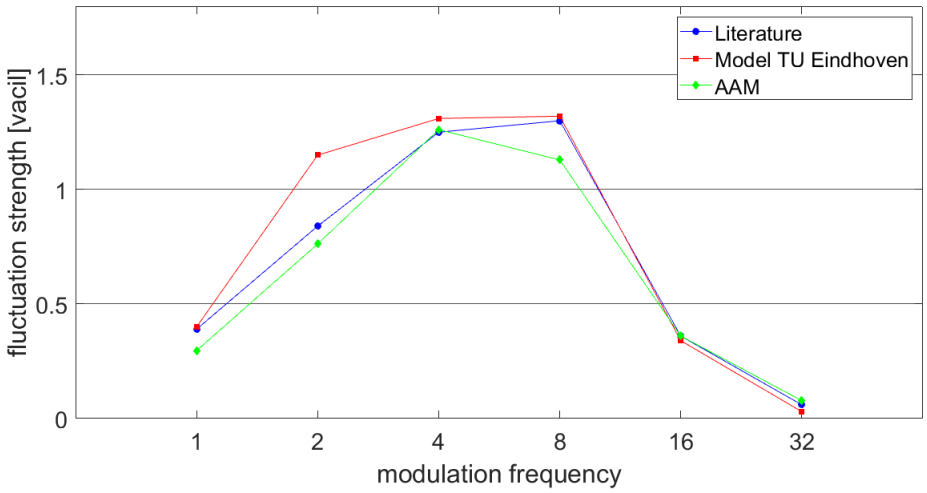


Figure 4.15: Fluctuation strength vs modulation frequency for AM tones

The stimuli had a center frequency of 1000 Hz, SPL of 70 dB and a modulation index of 1. It can be observed that the AAM results correspond very well with results from literature. The model of TU Eindhoven shows a large discrepancy for $f_{mod} = 2$ Hz. Furthermore, it can be noticed that the highest value is obtained at $f_{mod} = 4$ Hz for the AAM results, while the maximum value occurs at a modulation frequency of 8 Hz for the other two curves. This may be ascribed to the fact that the filters accounting for the modulation frequency dependency are more accurate in the AAM, as has been described in section 3.2.2.

In figure 4.16 the fluctuation strength is plotted against modulation frequency for FM tones. The stimuli had a center frequency of 1000 Hz, SPL of 70 dB and frequency deviation of 700 Hz.

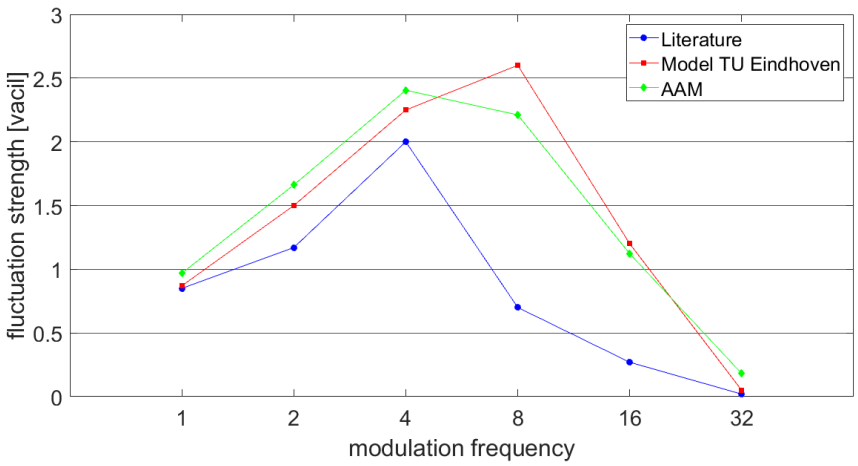


Figure 4.16: Fluctuation strength vs modulation frequency for FM tones

For both the model of TU Eindhoven and the model implemented in the AAM large differences can be seen at high modulation frequencies with respect to the results from literature. This is as expected since the model is derived from the roughness method of Daniel and Weber which is optimized for AM tones. However, the peak for the AAM results do occur at a modulation frequency of 4 Hz, which is also the case in the results found from literature.

In figure 4.17 the fluctuation strength is plotted against modulation frequency for AM broadband noise. The noise had a bandwidth of 16 kHz, SPL of 60 dB and modulation index of 1.

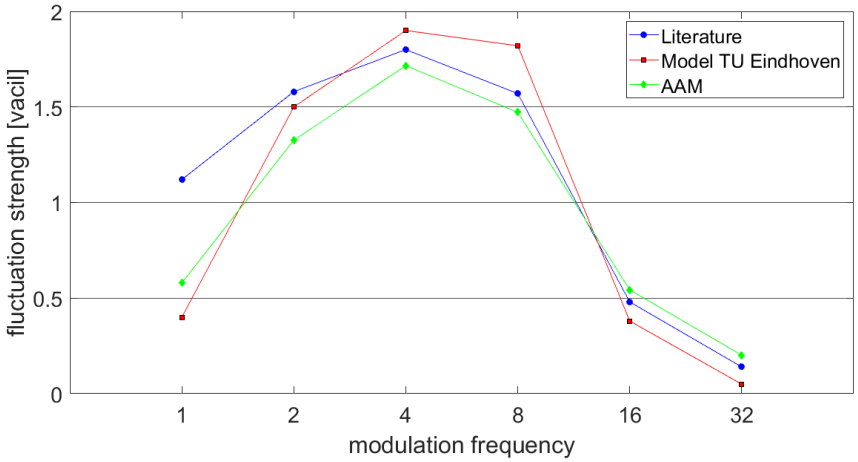


Figure 4.17: Fluctuation strength vs modulation frequency for AM broadband noise

For low modulation frequencies, the AAM model and TU Eindhoven model underestimate the fluctuation strength. At higher modulation frequencies, a better agreement can be observed between literature and the models. Here the peak value occurs at a modulation frequency of 4 Hz for all three curves.

As with roughness, the validation of fluctuation strength is also performed with Artemis by making use of "everyday" sounds. The same sounds are included again in the analysis. Fluctuation strength exceeded 5% of the time is calculated with both the AAM and Artemis and a linear regression is drawn through these points, as can be seen in figure 4.18. Now a slightly lower R^2-value can be seen than for the roughness analysis. This is because the fluctuation strength metrics are derived from the roughness metric and thus are derived from different methods. Furthermore, it was also found that the settling phase is very long (approximately 5 seconds) for Artemis which also distorts the results. This is due to the transient effect of digital filters at the beginning of the analysis [46].

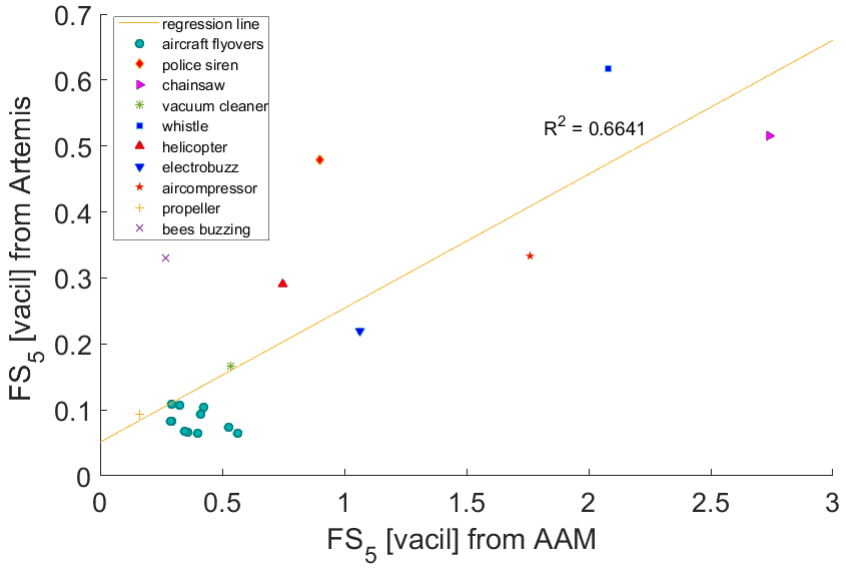


Figure 4.18: Correlation between results generated with the AAM and Artemis

The validation of fluctuation strength has been done using data found in literature and using Artemis. It is confirmed that in the Matlab implementation the fluctuation strength values peak around a modulation frequency of 4 Hz. The results generated with Artemis showed reasonable agreement with results generated with the AAM.

5

APPLICATION TO MEASURED SOUNDS

In this chapter the AAM is applied to measured aircraft sounds. In figure 5.1 the flow diagram is illustrated for the AAM with all its capabilities. The calculation starts with loading an audio file. After the initialization of variables, the SPL is computed. This is followed by the calculation of the conventional metrics and the five SQ metrics. The SQ metrics are used to calculate a psychoacoustic annoyance value. At the end some figures, for example spectrograms, are plotted and all the relevant data is written in a dat/excel file.

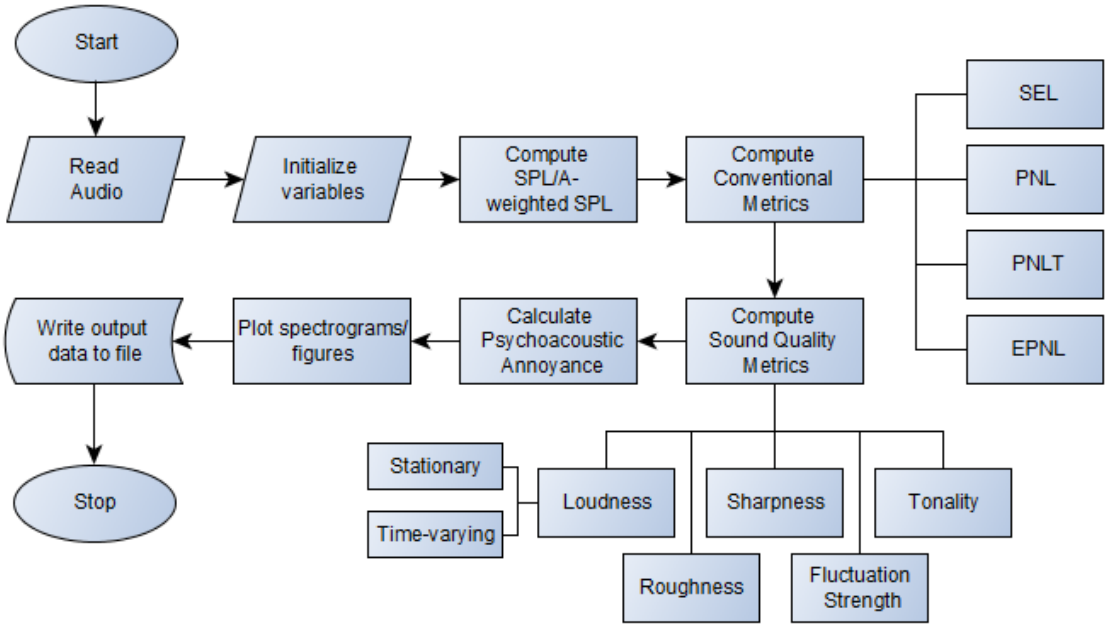


Figure 5.1: Flow diagram AAM

The data used consisted of aircraft flyovers measured at Schiphol airport. A total of 255 aircraft flyovers were used. The aircraft sounds measured were for 26 different aircraft, which are shown in table 5.1.

Table 5.1: Aircraft types and corresponding number of aircraft flyovers

Aircraft	Number of flyovers	Aircraft	Number of flyovers
Boeing 737-300	4	Airbus A300-600	1
Boeing 737-400	2	Airbus A319	7
Boeing 737-500	3	Airbus A320	14
Boeing 737-700	25	Airbus A321	6
Boeing 737-800	72	Airbus A330-200	1
Boeing 737-900	9	Airbus A380-800	3
Boeing 747-400	3	Embraer E145	2
Boeing 747-8	2	Embraer E175	15
Boeing 767-300	1	Embraer E190	32
Boeing 777-200LR	8	Fokker F70	27
Boeing 777-300	4	Bombardier CRJ900	2
Boeing 787-800	3	Bombardier CRJ1000	1
Boeing 787-900	5	Avro RJ-85	3

Results for each aircraft flyover were generated with the AAM. The five sound quality metrics are the relevant outputs for this study. To indicate the degree of dispersion and the skewness of the generated results, boxplots are given for each sound quality metric in figures 5.2 - 5.6. On the abscissa the aircraft type is given.
Outliers are shown as individual points (+). For an aircraft type with only one flyover measurement, like the Boeing 767-300, the boxplot is a single line. For some aircraft, like the Boeing 747-8, the boxplot is comparatively tall, indicating high variety in the sound quality metrics. For other aircraft, like the Embraer E175, the boxplot is comparatively short. When observing the loudness boxplots (figure 5.2), it can be observed that the boxplots for bigger aircraft, like the Boeing 747 or Airbus A380, is higher than other types of aircraft. For almost every aircraft type, a lot of variation can be seen in the tonality boxplots (figure 5.3) and sharpness boxplots (figure 5.5) . The median values for roughness (figure 5.4) and fluctuation strength (figure 5.6) are fairly identical for each aircraft type.

In section 5.1 a design parameter analysis is performed using the results from the AAM. It will be investigated how different design parameters have an impact on different sound characteristics such as tonality, loudness etc.. As stated in chapter 2, the sound quality metrics can be combined in a psychoacoustic annoyance model. This model will be compared with the traditional certification metric EPNL using subjective evaluations in section 5.2.

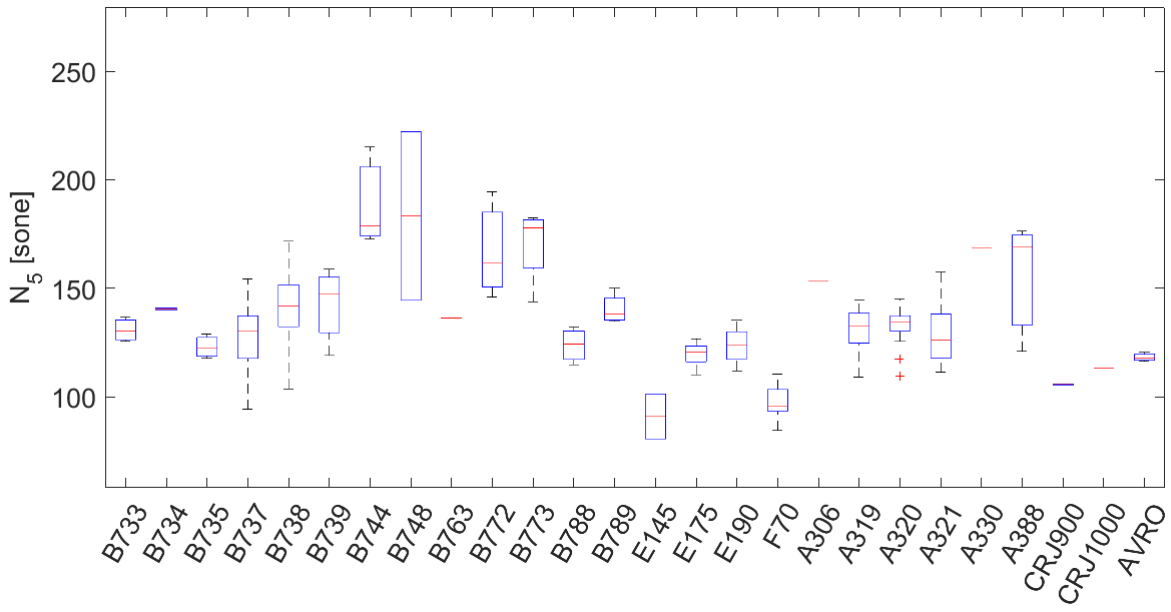


Figure 5.2: Boxplot of loudness exceeded 5% of the time

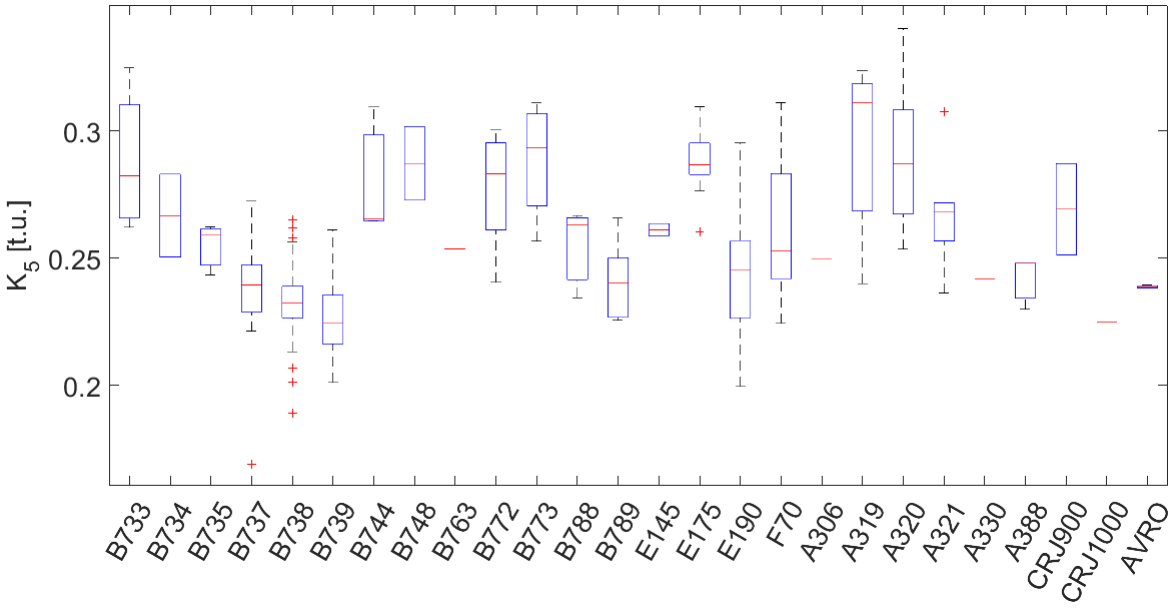


Figure 5.3: Boxplot of tonality exceeded 5% of the time

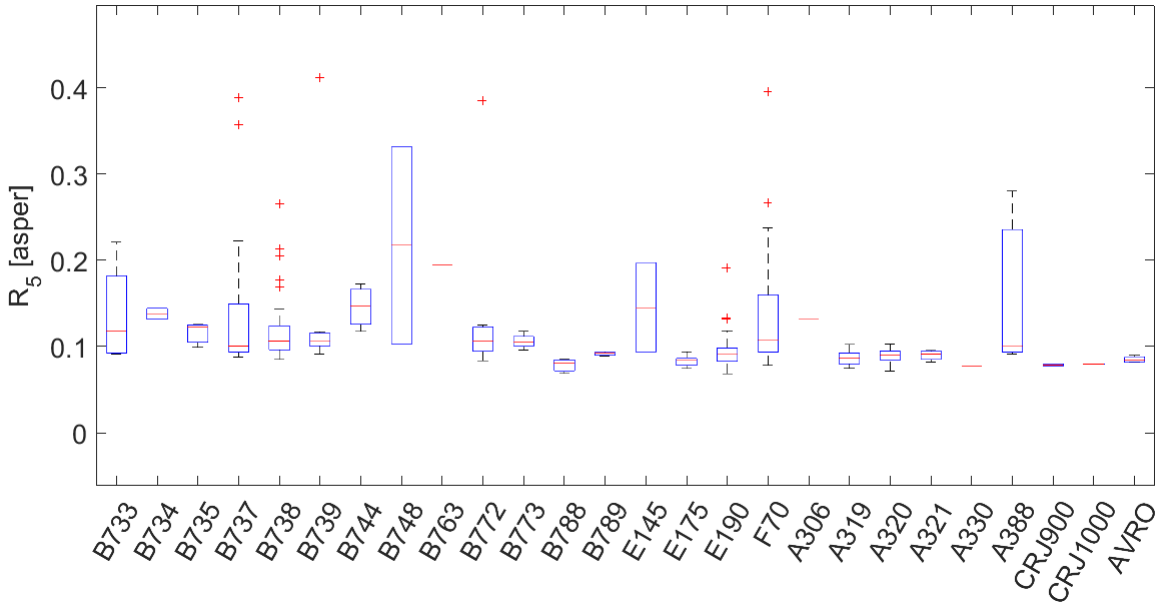


Figure 5.4: Boxplot of roughness exceeded 5% of the time

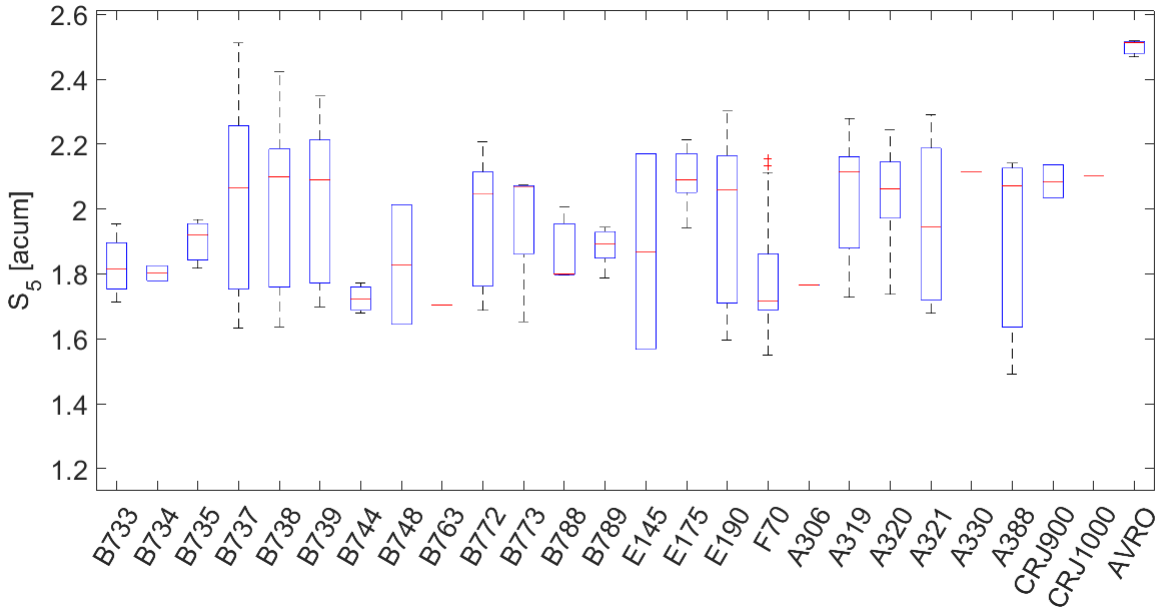


Figure 5.5: Boxplot of sharpness exceeded 5% of the time

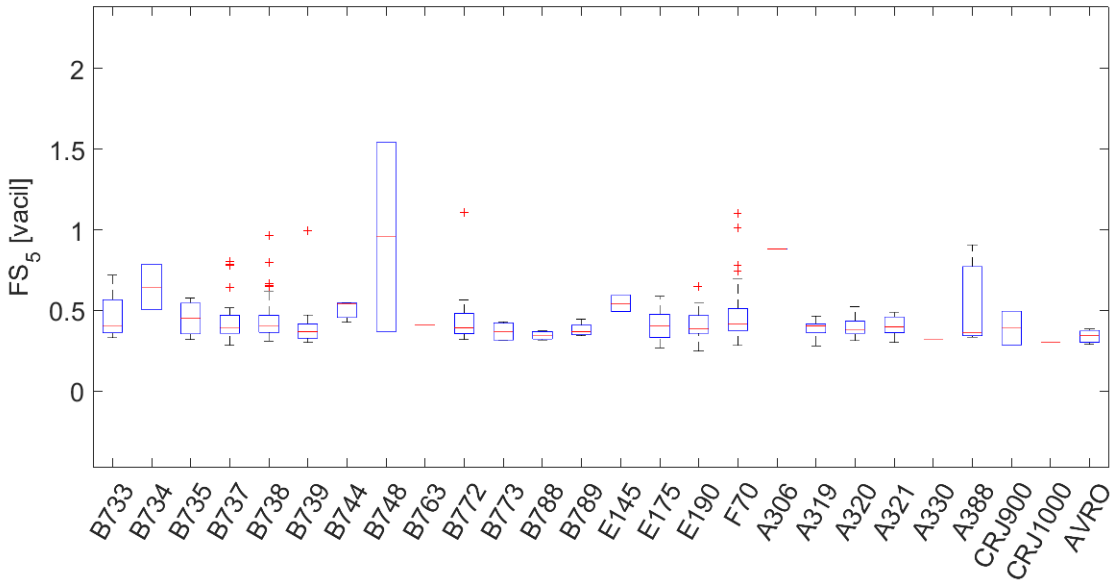


Figure 5.6: Boxplot of fluctuation strength exceeded 5% of the time

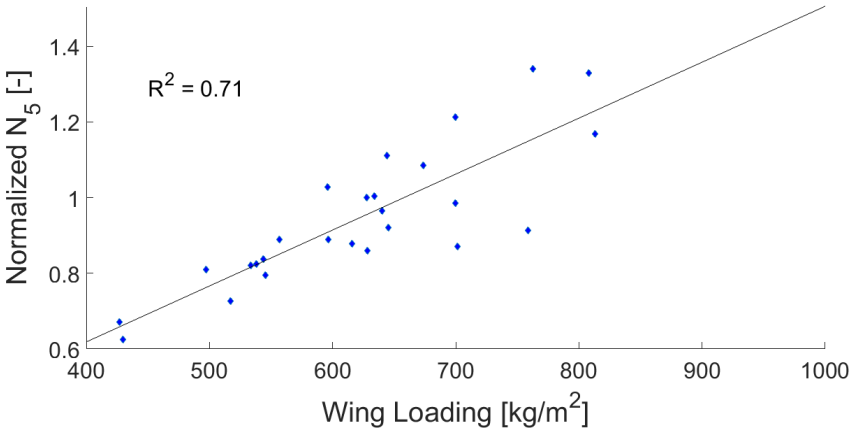
5.1. DESIGN PARAMETER ANALYSIS

In this section some design variables will be analyzed using the measured aircraft flyovers. It is attempted to find relations between these design variables and the five sound quality metrics. Additionally, it is investigated what kind of impact the engine location has on the sound quality metrics. For each aircraft type, all flyovers for that aircraft type were used and the average was calculated for each sound quality metric. The calculated results for each aircraft type are given in Appendix B. Finally, five helicopter flyovers are analyzed to investigate whether some differences with respect to turbofan aircraft can be observed.

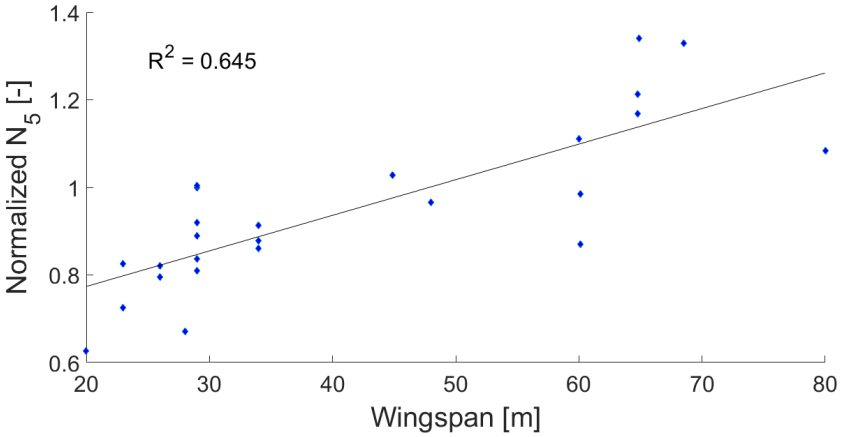
5.1.1. WING

The two variables investigated were the wingspan b and the wing loading W/S. These parameters change the aircraft's flight path along with the airframe geometry and thus change the airframe noise. The wingspan varied from 20 meters to 80 meters, while the wing loading varied between 426.90 kg/m^2 and 813.03 kg/m^2. A relation was sought between the wingspan/wingloading and the five sound quality metrics by means of a linear regression through the data points. The correlation between wingspan/wing loading and N_5 was the only significant correlation observed. These correlations can be seen in figure 5.7. The loudness values on the ordinate are normalized with respect to the Boeing 737-800. The P-values for the correlations of loudness versus wing loading and wing span are 6.72e-08 and 7.96e-07 respectively. This means that these correlations are statistically significant. Apart from the loudness, the other sound quality metrics showed little to no correlation. A higher wingspan translates into a higher loudness value, since more airframe noise is being

produced. This also means that an aircraft with a higher wingspan causes a more annoying aircraft since the loudness has the largest contribution on the annoyance experienced by residents [47], [7]. An increase in wing loading increases the loudness. It must be noted that a higher wing loading does not necessarily mean a higher weight or a lower wing size. If the weight would be kept constant and the wing size is increased, you would have a lower wing loading and expect a louder aircraft. However, in the aircraft analyzed in this section the higher wing loadings correspond to larger aircraft, thus larger wings which lead to a higher loudness values.



(a) Linear regression between wing loading and normalized N_5



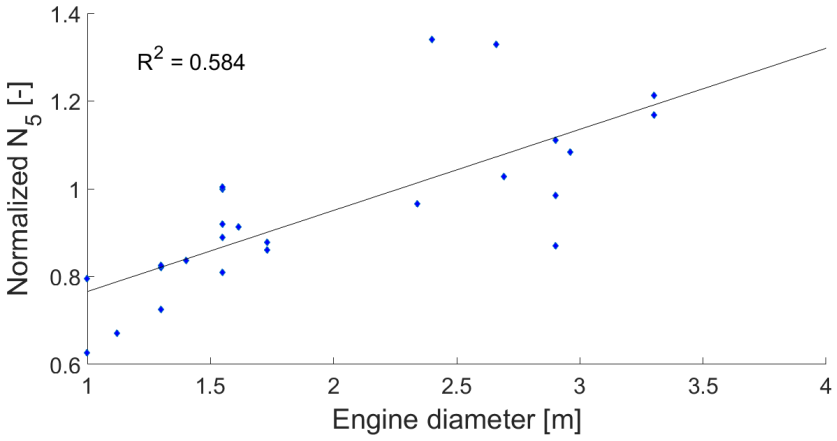
(b) Linear regression between wingspan and normalized N_5

Figure 5.7: Wing parameters analysis

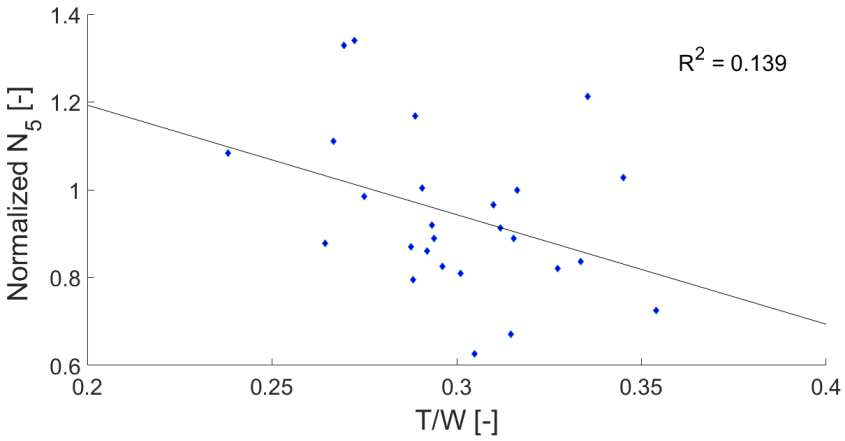
5.1.2. ENGINE

Engine noise is one of the main contributors to aircraft noise and hence the engine parameters are interesting to analyze. The three parameters investigated in this section were the following: engine diameter D_e, engine bypass ratio BPR and thrust-to-weight ratio T/W. The engine diameter varied between 1 meter and 3.3 meters. Again, besides the

N_5, there is no correlation between engine diameter and the other sound quality metrics. The engine diameter and loudness correlation shows a R^2-value of 0.58. The P-value for this regression analysis is 5.59e-06, and thus this regression is statistically significant. This does not necessarily mean that a larger engine creates more engine noise. Larger engines are mounted on bigger aircraft, thus with a larger body, which could be the source for the increase in airframe noise and thus loudness, as found previously by the strong relation between wingspan and N_5. Larger turbofans are in fact known to reduce jet noise due to the higher bypass ratio. This is somewhat illustrated when looking at the correlation between T/W and loudness. It can be seen that there is a very weak, but more importantly, a negative correlation. Thus aircraft which produce more thrust for the same weight are not louder. The P-value for this regression was 0.06. All other correlations of T/W with the sound quality metrics were weak. For the BPR all correlations were weak, with the highest R^2-value being 0.23 for loudness. All other correlations are given in Appendix C.



(a) Linear regression between engine diameter and normalized N_5



(b) Linear regression between T/W and normalized N_5

Figure 5.8: Engine parameters analysis

It is also investigated what impact the engine location has on the sound quality metrics. For that purpose, aft-mounted engines were compared with wing-mounted engines. Since aft-mounted engines are used on smaller aircraft, only the small aircraft with wing-mounted engines were used in this analysis. The results are given in table 5.2.

Table 5.2: Values of the SQ metrics for aft-mounted and wing-mounted aircraft analysis

	Aft-mounted	**Wing-mounted**
N_5 [sone]	98.7877	119.5368
K_5 [t.u]	0.2593	0.2801
R_5 [asper]	0.1603	0.0834
S_5 [acum]	1.8407	2.1584
FS_5 [vacil]	0.4804	0.3925

The engines mounted on the tail have a lower loudness value. This may be due to some noise being shielded by the fuselage. Also a lower tonality and lower sharpness value can be observed. This indicates that tones are more effectively masked with aft-mounted engines and that the high-frequency content of the noise is decreased. The aircraft with aft-mounted engines show higher values for roughness and fluctuation strength.

5.1.3. HELICOPTER SOUNDS ANALYSIS

In this section five helicopter sounds are analyzed in order to see the differences between helicopters and turbofan aircraft. The results of the analysis are given in table 5.3.

Table 5.3: Results of the helicopter sounds analysis

Measurement	N_5 **[sone]**	K_5 **[t.u]**	R_5 **[asper]**	S_5 **[acum]**	FS_5 **[vacil]**	PA_{mod}	**EPNL**
1	22.7690	0.3477	0.7344	1.0386	0.7872	44.7421	87.2361
2	12.3161	0.2313	0.2946	0.4794	0.2597	20.9628	87.2985
3	31.1739	0.2721	1.0323	0.7737	0.4168	34.9796	91.2772
4	21.737	0.3004	0.8363	1.3022	1.0362	31.3592	85.3859
5	40.4481	0.398	0.4402	1.1224	0.7245	55.2383	90.5948

The most notable difference between helicopters and turbofan aircraft is the roughness value. The helicopter sounds are much rougher due to the buzzing sound a helicopter rotor makes. The loudness values are lower, since the helicopters were flying quite a distance away from the mics. The tonality and fluctuation strength values are of the same order as for the turbofan aircraft. The lower sharpness values indicate that there is less high-frequency content which can be explained by the fact that there is no fan which generates high-frequency tones. Additionaly, it can be seen that the EPNL and PA_{mod} values are not

consistent. Measurement 4 has the lowest EPNL, while the PA_{mod} indicates measurement 2 as least annoying. Furthermore, measurement 3 has the highest EPNL, while the PA_{mod} indicates measurement 5 as most annoying. The PA_{mod} metric was created using aircraft noises [12], and is thus not suitable for assessment of helicopter noise. A psychoacoustic annoyance metric for helicopters should have a larger contribution of roughness and smaller contribution of sharpness.

5.2. COMPARISON WITH SUBJECTIVE EVALUATIONS

In figure 5.9 a linear regression is drawn for the 26 aircraft types from the previous section for EPNL and PA_{mod}. It can be observed that there is a high correlation between PA_{mod} and EPNL. This indicates that in almost 90% of the cases an increase in PA_{mod} is accompanied by an increase in EPNL. However, there are some cases in which the two metrics contradict each other.

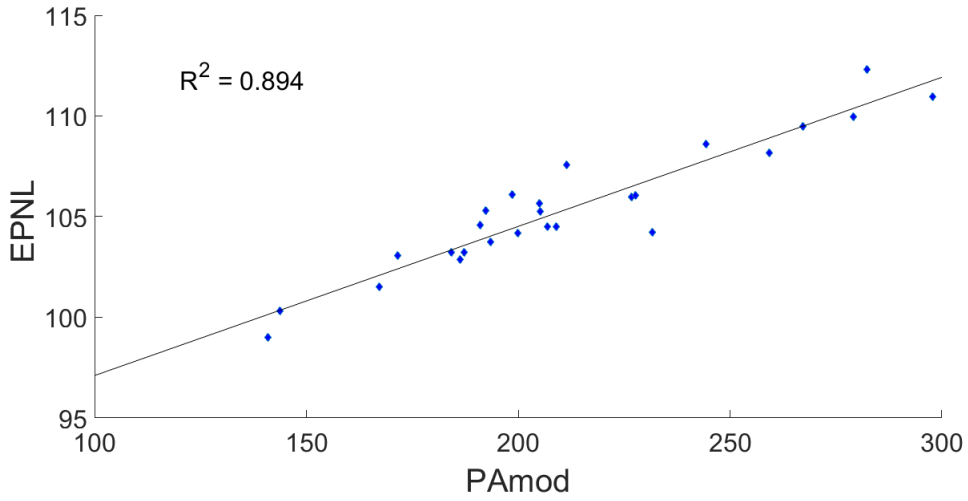


Figure 5.9: Correlation between PA_{mod} and EPNL

It is clear that the psychoacoustic annoyance model captures the various sound characteristics in a more sophisticated way than the traditional certification metrics such as EPNL. The question then arises whether the psychoacoustic annoyance value is a better predictor of annoyance and thus a better certification metric. For this purpose, subjective listening tests were conducted. First, the test setup is given. Secondly, the results of the experiment are discussed.

5.2.1. TEST SET-UP

Listening Arrangement

The tests were conducted in the anechoic chamber of the Faculty of Applied Sciences of the TU Delft. Twenty subjects (13 male/7 female, aged from 21 to 61) participated in the exper-

iment. The test-setup is schematically drawn in figure 5.10. The sound was played through two loudspeakers. The positioning of sound sources and listeners was done according to the recommendations given in [48]. Recommendations concern for example the distance of speakers from the walls, the distance between the speakers etc.

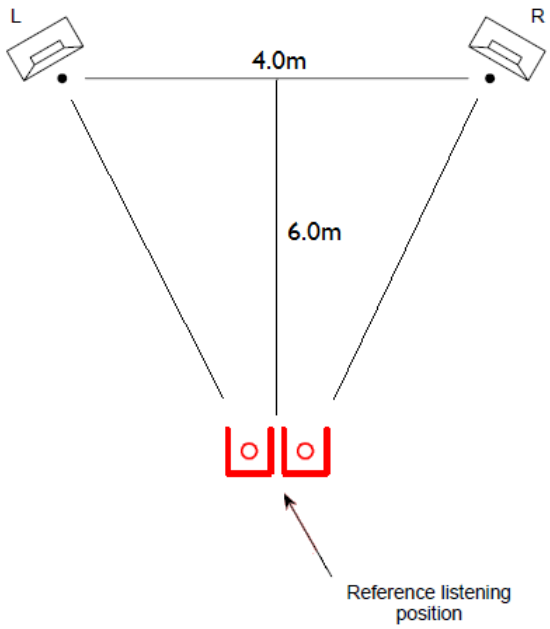


Figure 5.10: Test set-up

Method

The experiment consisted of two parts. The method used in the first part of the listening tests was the two alternative forced choice method. Two sounds were presented to the subjects and the subjects were asked to choose which one is more annoying. The sounds were selected in such a way that the EPNL and PA_{mod} values contradicted each other. So going from sound A to sound B, the EPNL increased while the PA_{mod} value decreased or vice versa. The direct scaling method was used in the second part of the listening tests. Ten sounds were presented to the subjects which they had to rate from 0-100 according to the scale given below.

Figure 5.11: Annoyance scale used in part 2 of the listening test

Stimuli

Part 1 of the listening tests consisted of 10 comparisons, so 20 sounds. In the second part another ten sounds were presented, amounting to a total of 30 sounds. The sounds were selected from the dataset of 255 aircraft flyovers introduced in the previous section.

5.2.2. RESULTS

The results of the two alternative forced choice part are given in table 5.4. Ten sets of sounds can be seen with their PA_{mod} and EPNL values. For each subject the cell is colored for the sound which the subject perceived to be more annoying. Additionally, the percentage of people who indicated that sound to be more annoying is given.

For part 2 of the test, 20 subjects gave ratings to 10 sounds, amounting to a total of 200 ratings. For each sound the average rating was calculated and a linear regression was drawn with EPNL and PA_{mod}. This can be seen in figures 5.12 and 5.13. Additionally, correlations of the five sound quality metrics with the average annoyance rating is also given (figures 5.14 - 5.18). The ratings given by each subject are given in Appendix D.

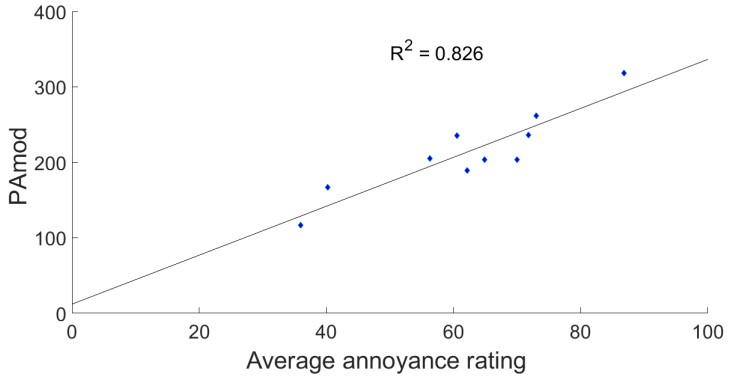


Figure 5.12: Correlation between subjective annoyance ratings and PA_{mod}

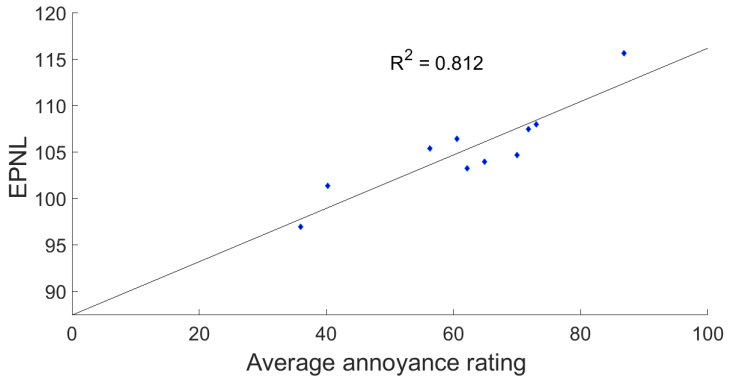


Figure 5.13: Correlation between subjective annoyance ratings and EPNL

Table 5.4: Results two alternative forced choice test

		PA_{mod}	EPNL	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15	Subject 16	Subject 17	Subject 18	Subject 19	Subject 20	%
Set 1	Sound A	300.79	107.86				■				■							■	■		■		■	30
	Sound B	282.43	109.17	■	■	■		■	■	■		■	■	■	■	■	■			■		■		70
Set 2	Sound A	277.51	107.96																			■	■	10
	Sound B	273.95	110.18	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■			90
Set 3	Sound A	125.98	99.11	■		■		■	■				■		■		■			■	■			45
	Sound B	138.20	98.08		■		■			■	■	■		■		■		■	■			■	■	55
Set 4	Sound A	149.23	102.60									■								■	■	■		20
	Sound B	157.24	101.81	■	■	■	■	■	■	■	■		■	■	■	■	■	■	■				■	80
Set 5	Sound A	152.96	102.41				■		■		■			■		■	■	■			■	■	■	50
	Sound B	167.49	101.29	■	■	■		■		■		■	■		■				■	■				50
Set 6	Sound A	174.12	104.09	■	■	■									■	■		■	■				■	40
	Sound B	186.29	102.88				■	■	■	■	■	■	■	■			■			■	■	■		60
Set 7	Sound A	190.66	105.15	■			■	■												■		■		25
	Sound B	203.38	104.16		■	■			■	■	■	■	■	■	■	■	■	■	■		■		■	75
Set 8	Sound A	206.53	106.04	■	■	■	■	■		■	■	■		■	■						■	■		60
	Sound B	220.84	104.58						■				■			■	■	■	■	■			■	40
Set 9	Sound A	224.47	107.67	■	■			■	■		■	■	■		■	■					■	■	■	60
	Sound B	240.08	106.48			■	■			■				■			■	■	■	■				40
Set 10	Sound A	318.02	115.63	■	■	■	■	■	■		■		■	■	■	■	■			■		■		70
	Sound B	371.21	114.28							■		■						■	■		■		■	30

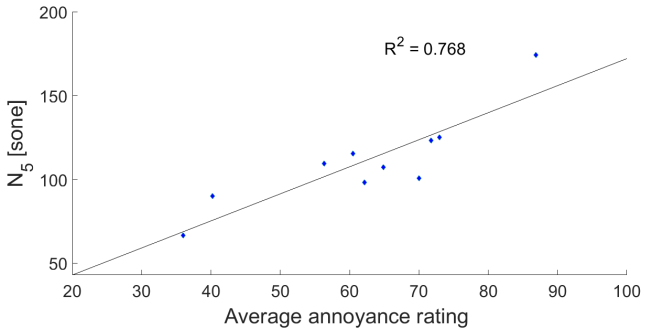


Figure 5.14: Correlation of subjective annoyance rating with loudness (P-value = 0.000875)

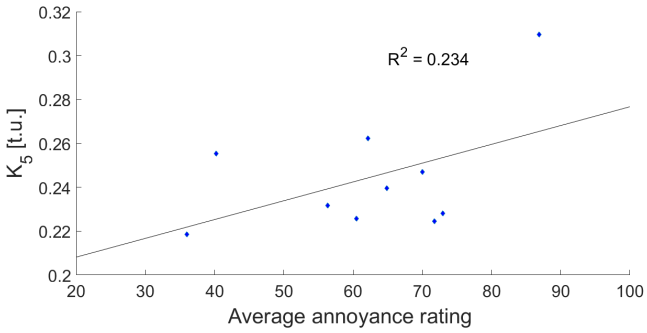


Figure 5.15: Correlation of subjective annoyance rating with tonality (P-value = 0.157)

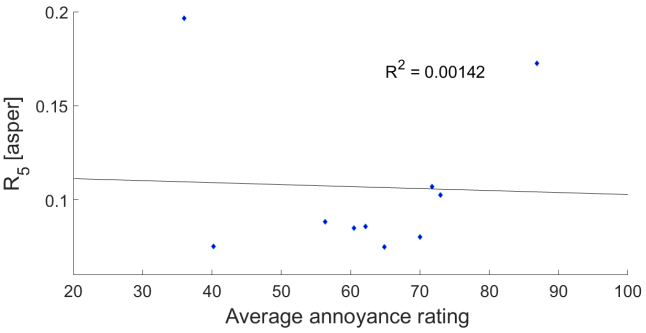


Figure 5.16: Correlation of subjective annoyance rating with roughness (P-value = 0.918)

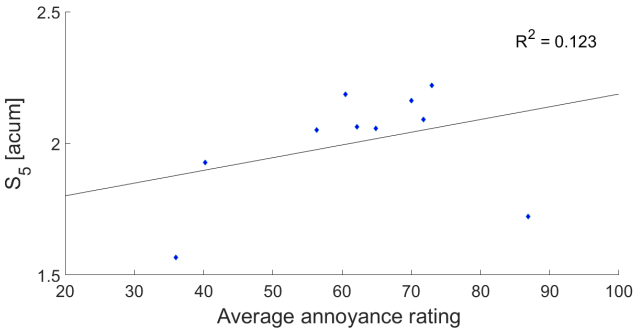


Figure 5.17: Correlation of subjective annoyance rating with sharpness (P-value = 0.32)

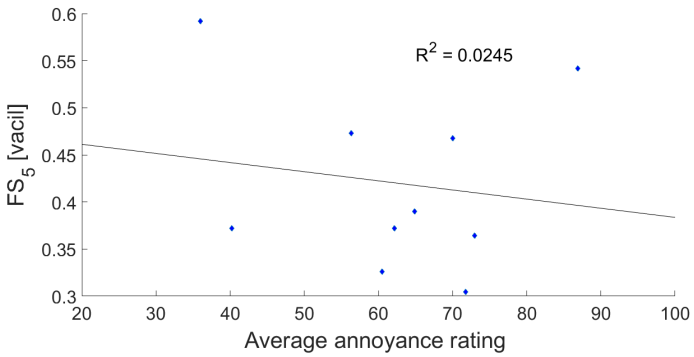


Figure 5.18: Correlation of subjective annoyance rating with fluctuation strength (P-value = 0.666)

5.2.3. DISCUSSION

Looking at the results presented in table 5.4, no clear conclusion can be drawn about which of the two (EPNL or PA_{mod}) is better in predicting the annoyance experienced by the subjects. In some cases, e.g. set 2, most subjects indicated the sound with the higher EPNL as more annoying. In other cases (set 4) most subjects indicated the sounds with the higher PA_{mod} value as more annoying. And in some other cases (set 5) the sounds were indicated to be annoying just as often. This indicates that even though the pyschoacoustic annoy-

ance model incorporates the sound characteristics in a much more sophisticated way than EPNL, it is not necessarily a better annoyance predictor. This can also be observed by looking at figures 5.12 and 5.13. The R^2-value for PA_{mod} is only sightly higher than for EPNL. The P-values were 0.000371 and 0.000273 for the EPNL and PAmod correlation respectively, and thus these correlations are statistically significant. The subjective results are presented as average annoyance ratings, varying from a minimum observed annoyance of 36 and a maximum observed annoyance of close to 87, for the 10 aircraft flyovers.

A possible reason for the PA_{mod} not correctly predicting the annoyance experienced by the subjects may be the fact that it uses the "values exceeded 5% of the time" for each metric. When using these values, the duration of the flyover becomes important when comparing two sounds. A flyover with a longer duration (and thus a longer period where the aircraft is not heard) will lead to lower 'exceeded 5% of the time' values and thus a lower PA_{mod} value. The duration of the aircraft flyovers used in this listening tests varied between 17 seconds and 24 seconds.

This dependency on duration was confirmed when investigating the sounds given in set 2. During the listening test, 90 % of the subjects indicated sound B to be more annoying. However, sound B has a lower PA_{mod} value than sound A and it was suspected that the longer duration of sound B could be a reason for this. Sound A of set 2 has a duration of 17 seconds and sound B has a duration of 20 seconds. Sound B was shortened to 17 seconds and loaded once again in the AAM. This time it gave a PA_{mod} value of 291.08, which is higher than the PA_{mod} value for sound A and thus consistent with the annoyance experienced by the subjects. So in order to make a good comparison in terms of PA_{mod} between two sounds, the duration of the sounds have to be the same. Another possibility is to not take the entire duration of the sound file into account while computing the PA_{mod} value and adopt an approach similar to the duration correction in the EPNL metric. For example, only the ten seconds around the maximum loudness value could be used in the calculation of the psychoacoustic annoyance.

From part 2 of the analysis, both the PA_{mod} and EPNL metric were seen to show high correlation with subjective annoyance ratings. Both the EPNL and PA_{mod} correlation show a coefficient of determination of around 0.8. Even though the correlations are high, there is some room for improvement. More research into PA_{mod} can potentially improve correlations with subjective results. In this regard, the effect and weightage of tonality need further investigation, to reduce the dominant influence of loudness in the predicted annoyance metrics. The dominant influence of loudness can be illustrated when looking at

set 10 of part 1 of the listening test. Almost all subjects indicated the sound with the higher EPNL value to be more annoying. This is possibly due to the piercing tone that can be heard in that sound. Still, the PA_{mod} value for that sound was significantly lower than the other sound. The values for the various sound quality metrics are given in table 5.5.

Table 5.5: SQ metrics for set 10 of part 1 of the listening test

Set 10	N_5 **[sone]**	K_5 **[t.u]**	R_5 **[asper]**	S_5 **[acum]**	FS_5 **[vacil]**	PA_{mod}	**EPNL**
Sound A	174.1974	0.3097	0.1726	1.7222	0.5416	318.0170	115.6283
Sound B	189.2363	0.3017	0.1032	2.0136	0.3693	371.2072	114.2795

It can be seen that the only large difference between the two sounds is the value of N_5. This, together with the fact that most subjects indicated sound A to be more annoying, shows that the loudness contribution in the calculation of PA_{mod} might be excessive. It also shows that the tonality contribution to the annoyance experienced by people is underrated in the PA_{mod} calculation. It might be unnecessary to include loudness in the tonal weighting term w_{Ton} (see equation 2.31), when loudness is already incorporated in the calculation of Aures' tonality. Another point to be noted is that the sharpness has a relatively large influence on the PA_{mod} calculation, even though it is the fourth important metric for aircraft sound [7],[12]. For sound A, the sharpness is completely omitted since it is lower than 1.75 acum. This suggests that the sharpness contribution in the PA_{mod} calculation should be carefully investigated.

Analyzing the correlations of the individual metrics with the annoyance ratings, it can be seen that the only statistically significant correlation is that of loudness. It also shows the highest value for the coefficient of determination. In terms of the coefficient of determination, tonality comes at the second place, indicating that it is indeed the second most important sound quality metric for aircraft noise assessment. Interestingly enough, not roughness but sharpness has the third highest value for R^2. As is expected, fluctuation strength shows little correlation with the average annoyance ratings. Even though the PA_{mod} showed high correlation with the subjective annoyance ratings, the findings mentioned in this section illustrate that it is possible and necessary to optimize the PA_{mod} metric.

Like all audio quality listening tests, this listening test is not free of biases. An example of a bias type encountered is the "recency effect" [49]. This effect means that the results of the audio quality assessment is biased towards that part of the test that was heard most recently. In other words, subjects would tend to choose sound B as more annoying since this was the sound which they heard more recently than sound A. It was tried to reduce

this bias by changing the order in which the sounds were played for every time the listening test was performed. Other biases typically encountered during listening tests are bias due to listener expectations and the contraction bias. The first bias was tried to reduce by using as many subjects as possible with different backgrounds. Doing the two alternative forced choice part first, helps familiarizing the listeners with the stimuli under assessment and thus reducing the contraction bias in the direct scaling part of the test. While taking these kind of measures helps reducing the biases, there will still be some bias and thus it is important to be aware of the biases encountered in audio listening tests.

6

CONCLUSION & RECOMMENDATIONS

This master thesis had the objective of completing the AAM by implementing the SQ metrics tonality, roughness and fluctuation strength. The metrics loudness and sharpness had already been incorporated in the AAM. For each metric, different methods exist and thus a decision had to be made on which method to use. It was decided to implement Aures' tonality method since many studies highlighted its suitability for aircraft noise assessment. In Aures' tonality method a tonal weighting factor is calculated for tonality after taking into account the tonal bandwidth, the tone frequency and the tone emergence level. Parallel to the tonal weighting factor, a loudness weighting factor is calculated. Finally, both the tonal and loudness weightings are combined to find a value for tonality.

For roughness, the method of Daniel & Weber was used due to its proven ability to capture changes in buzzsaw noise. The total roughness is computed from the specific roughness in each critical band. The specific roughness is calculated from the modulation depth. Weighting functions accounting for the dependency on modulation frequency, and calibration factors accounting for the dependence on carrier frequency are used. The fluctuation strength implementation was derived from the roughness implementation since these two metrics are similar to each other. The three metrics implemented were validated using publicly available data and the commercial audio quality assessment software Artemis.

Subsequently, the AAM was applied to measured aircraft flyover sounds. The first part of the application consisted of a design parameter analysis. It was attempted to find relations between the design variables and the sound quality metrics. The only significant correlations observed were those of the wingspan, wing loading and engine diameter with

loudness. It must be noted that these correlations could also indicate an indirect relation between the variables. For example, a larger engine diameter does not automatically translate into a higher loudness value, but larger engines are mounted on bigger aircraft and the larger body causes an increase in airframe noise and thus loudness. Aircraft with aft-mounted engines were shown to have a lower loudness, tonality and sharpness than aircraft with wing-mounted engines. Noise shielding due to the fuselage might be a reason for the lower values, but further research is needed to find out what the exact effects of engine position on sound quality are. Helicopters show a much higher value for roughness than turbofan aircraft due to the buzzing sound a helicopter rotor makes. This indicates that roughness is an important metric for aircraft with engine architectures such as the open rotor.

To investigate whether the psychoacoustic annoyance value is a better annoyance predictor than EPNL, listening tests were set up. From the results of these tests it was found that the metrics did not outperform each other in terms of annoyance assessment. Both the EPNL and PA_{mod} showed a similar correlation with the subjective annoyance ratings. The two alternative forced choice test was performed to investigate two sounds which contradicted each other in terms of PA_{mod} and EPNL. In some cases, the sound with the higher EPNL value was found to be more annoying, while in other cases the sound with the higher PA_{mod} value was found to be more annoying.

Some recommendations for future research can be stated with regard to the data. The design parameter analysis conducted in this research made use of 255 aircraft flyovers of 26 different aircraft types. The number of flyovers for each type of aircraft ranged from 1 to 72. For this type of analysis it is recommended to use a minimum of 10 aircraft flyovers for each aircraft type in order to have a more fair comparison. By taking the average of more than one aircraft flyover, the effect of disturbances such as wind will be minimized. Furthermore, the aircraft flyovers varied in duration between 15-24 seconds. The duration has an influence on the PA_{mod} value, since it uses the "exceeded 5% of the time" values. It is therefore recommended to use flyovers of the same duration when comparing two sounds. All aircraft flyover measurements used in the design parameter analysis and listening tests were of aircraft approaching the runway. It might also be interesting to see how the sound quality metrics differ when aircraft flyover measurements of aircraft departing from the airport are used. Only turbofan aircraft were used in the development of the PA_{mod} metric. It is suggested to analyze propeller aircraft sounds with the AAM to develop a psychoacoustic annoyance model that is suitable for propeller aircraft. This will require subjective evaluations of propeller aircraft sounds.

It is also recommended to perform a design parameter analysis in a dedicated design environment using auralized sounds. By varying only one design parameter and keeping all other design parameters constant, direct relations between design parameters and sound quality metrics will become clear. The design parameter analysis conducted in this research was a bivariate analysis, i.e. an analysis to find out if there is a relationship between two variables. It is suggested to perform a multivariate analysis in which several design variables are examined at the same time. This gives a much more realistic picture than trying to explain the effect on a dependent variable by looking at only one single design variable. Moreover, it is recommended to perform an analysis using synthesized sounds with subtle changes in sound such as variations in buzzsaw noise intensity. This will clarify whether the implemented metrics are able to notice subtle changes in sound. It might also be interesting to implement other methods of the sound quality metrics to compare them with the ones implemented during this research.

Some deficiencies of the PA_{mod} metric came to light during the listening tests conducted in this research. One of the reasons the PA_{mod} did not correctly predict the annoyance all the time may be the varying duration of the aircraft flyovers. As stated above, for a valid comparison of two sounds in terms of PA_{mod}, the duration of the sounds has to be the same. It was also found that the loudness contribution to PA_{mod} might be too high. Also, the effect and weighting of tonality and sharpness need further investigation. More research into PA_{mod} can potentially improve correlations with subjective evaluations. An optimized psychoacoustic annoyance model should be validated with extensive listening tests.

BIBLIOGRAPHY

[1] M. Basner, W. Babisch, A. Davis, M. Brink, C. Clark, S. Janssen, and S. Stansfeld, *Auditory and non-auditory effects of noise on health,* The Lancet **383**, 1325 (2014).

[2] A.-S. Evrard, I. Khati, P. Champelovier, J. Lambert, and B. Laumon, *Health effects of aircraft noise near three french airports: results from pilot epidemiological study of the debats study.* in *AUN-2014'Airports in Urban Networks' European Conference* (2014) pp. 8–p.

[3] German Aviation Association (BDL), *Aircraft noise report,* (2015).

[4] D. Geng, K. Yi, C. Shang, J. Yang, and Y. He, *Application status of composite acoustic liner in aero-engine,* (2015).

[5] K. Zaman, J. Bridges, and D. Huff, *Evolution from tabs to chevron technology-a review,* International Journal of Aeroacoustics **10**, 685 (2011).

[6] International Civil Aircraft Organisation (ICAO), *Guidance on the balanced approach to aircraft noise management,* (2008).

[7] A. Sahai, *Consideration of Aircraft Noise Annoyance during Conceptual Aircraft Design,* Ph.D. thesis, RWTH Aachen University (2016).

[8] M. J. Lighthill, *On sound generated aerodynamically i. general theory,* Proc. R. Soc. Lond. A **211**, 564 (1952).

[9] A. K. Sahai, T. van Hemelen, and D. G. Simons, *Methodology for designing aircraft having optimal sound signatures,* The Journal of the Acoustical Society of America **141**, 3688 (2017).

[10] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models* (Springer Science & Business Media., 1999).

[11] International Civil Aviation Organization (ICAO), *Noise certification workshop,* (2006).

[12] S. More, *Aircraft Noise Characteristics and Metrics,* Ph.D. thesis, Purdue University (2010).

[13] A. Minard and P. Boussard, *Signal-based indicators for predicting the effect of audible tones in the aircraft sound at takeoff.* in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Vol. 253(2) (Institute of Noise Control Engineering., 2016) pp. 6410–6419.

[14] M. Snellen, R. Merino Martinez, and D. G. Simons, *Assessment of aircraft noise sources variability using an acoustic camera,* in *Proceedings of the 5th CEAS Air and Space Conference: Challenges in European Aerospace, Delft (The Nehterlands), 7-11 Sept. 2015* (Delft University of Technology, 2015).

[15] Z. Zhang and M. Shrestha, *Sound quality user-defined cursor reading control - tonality metric,* (2003).

[16] ISO 532-1:2017(E), *ISO 532-1 (2017). Acoustics – Method for calculating loudness – Zwicker method*, Standard (International Organization for Standardization, Geneva, CH, 2017).

[17] E. Terhardt, *Calculating virtual pitch,* Hearing research **1**, 155 (1979).

[18] M. Van Der Heijden and A. Kohlrausch, *Using an excitation-pattern model to predict auditory masking,* Hearing research **80**, 38 (1994).

[19] E. Zwicker, *Procedure for calculating loudness of temporally variable sounds.* The Journal of the Acoustical Society of America **62(3)**, 675 (1977).

[20] W. Aures, *Procedure for calculating the sensory euphony of arbitrary sound signals.* Acustica **59(2)**, 130 (1985).

[21] G. von Bismarck, *Sharpness as an attribute of the timbre of steady sounds.* Acta Acustica united with Acustica **30(3)**, 159 (1974).

[22] DIN 45692, *DIN 45692 (2009). Measurement technique for the simulation of the auditory sensation of sharpness*, Standard (German Institute for Standardization (DIN), Berlin, De, 2009).

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

[24] W. Aures, *Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale,* Acta Acustica united with Acustica **59(2)**, 130 (1985).

[25] S. H. Shin, J. G. Ih, T. Hashimoto, and S. Hatano, *Sound quality evaluation of the booming sensation for passenger cars.* Applied acoustics (2009).

[26] M. de Diego, A. Gonzalez, G. Pinero, M. Ferrer, and J. J. Garcia-Bonito, *Subjective evaluation of actively controlled interior car noise.* Acoustics, Speech, and Signal Processing **5**, 3225 (2001).

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

[28] A. Hastings and P. Davies, *An examination of Auress model of tonality.* Proc. Internoise **2** (2002).

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

[30] T. Pedersen, M. Sondergaard, and B. Andersen, *Objective method for assessing the audibility of tones in noise Joint Nordic Method,* (1999).

[31] R. Sottek, *Progress in calculating tonality of technical sounds.* in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Vol. 249(4) (Institute of Noise Control Engineering., 2014) pp. 3319–3327.

[32] S. More and P. Davies, *Issues in the evaluation of the tonality of nonstationary sounds containing timevarying harmonic complexes.* The Journal of the Acoustical Society of America **123(5)**, 3451 (2008).

[33] E. Terhardt, *On the perception of periodic sound fluctuation (roughness),* Acustica **30**, 201 (1974).

[34] U. Widmann and H. Fastl, *Calculating roughness using time-varying specific loudness spectra.* in *Proceedings Sound Quality Symposium* (1998) pp. 55–60.

[35] W. Aures, *A procedure for calculating auditory roughness,* Acustica **58**, 268 (1985).

[36] P. Daniel and R. Weber, *Psychoacoustic roughness: Implementation of an optimized model,* Acustica **83**, 113 (1997).

[37] V. Vencovský, *Roughness Prediction Based on a Model of Cochlear Hydrodynamics.* Archives of Acoustics **41(2)**, 189 (2016).

[38] A. Osses Vecchi, R. García León, and A. Kohlrausch, *Modelling the sensation of fluctuation strength,* in *Proceedings of Meetings on Acoustics 22ICA*, Vol. 28(1) (ASA, 2016) p. 050005.

[39] G. Q. Di, X. W. Chen, K. Song, B. Zhou, and C. M. Pei, *Improvement of Zwickers psychoacoustic annoyance model aiming at tonal noises.* Applied Acoustics **105**, 164 (2016).

[40] S. More and P. Davies, *An examination of the combined effects of loudness, tonalness, and roughness on annoyance ratings of aircraft noise,* The Journal of the Acoustical Society of America **127(3)**, 1740 (2010).

[41] L. Leylekian, M. Lebrun, and P. Lempereur, *An overview of aircraft noise reduction technologies.* AerospaceLab **6**, 1 (2014).

[42] E. Zwicker, H. Fastl, U. Widmann, K. Kurakata, S. Kuwano, and S. Namba, *Program for calculating loudness according to DIN 45631 (ISO 532B),* Journal of the Acoustical Society of Japan **12(1)**, 39 (1991).

[43] J. Schrader, *A matlab implementation of a model of auditory roughness,* (2002).

[44] C. Dendievel, C. Lambourg, and P. Boussard, *Comparison of listening test methods to evaluate tonality perception,* in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Vol. 253 (Institute of Noise Control Engineering, 2016) pp. 6462–6470.

[45] R. Sottek, F. Kamp, and A. Fiebig, *A hearing model approach to tonality.* in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Vol. 253(2) (Institute of Noise Control Engineering., 2013) pp. 6055–6055.

[46] HEAD Acoustics, *Application notes: Psychoacoustic analysis II,* (2016).

[47] B. Berglund, U. Berglund, and T. Lindvall, *Scaling loudness, noisiness, and annoyance of aircraft noise,* The Journal of the Acoustical Society of America **57**, 930 (1975).

[48] W. Hoeg, L. Christensen, and R. Walker, *Subjective assessment of audio quality-the means and methods in the ebu,* EBU Technical Review , 40 (1997).

[49] S. Zielinski, F. Rumsey, and S. Bech, *On some biases encountered in modern audio quality listening tests-a review,* Journal of the Audio Engineering Society **56**, 427 (2008).

A

Amplitude Modulation and Frequency Modulation

In figure A.1 a graphical illustration is given of amplitude-modulation. Suppose, $x(t)$ is a information signal. The second wave is the carrier wave with a certain carrier frequency f_{carr}. This results in a amplitude-modulated wave with a varying amplitude.

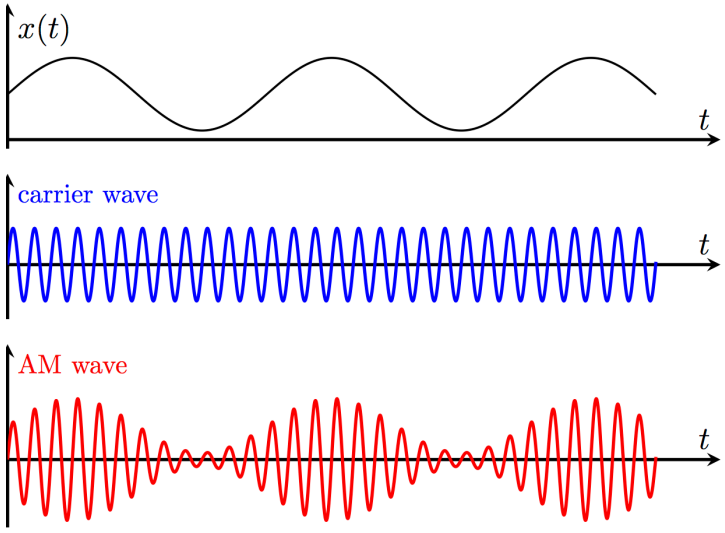


Figure A.1: Illustration of amplitude modulation [9]

Amplitude-modulation of a sine wave can be represented with equation A.1, in which m^* is the modulation depth, f_{mod} the modulation frequency and A_c the carrier amplitude.

$$y(t) = \left[1 + m^* \cdot sin(2\pi f_{mod} t)\right] \cdot A_c \cdot sin(2\pi f_{carr} t) \tag{A.1}$$

[9] Source of image: https://github.com/PetarV-/TikZ/tree/master/Amplitude%20modulation

In figure A.2 a graphical illustration is given of frequency-modulation. Modulating the carrier wave with the information signal results in a frequency-modulated wave with a varying instantaneous frequency.

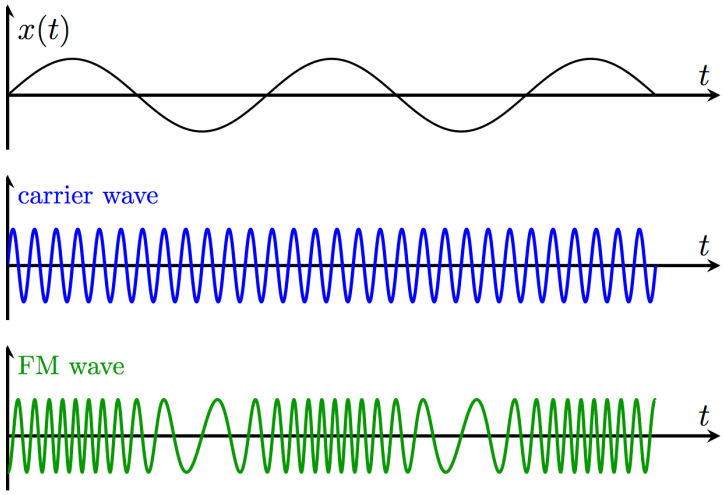


Figure A.2: Illustration of frequency modulation [10]

Frequency-modulation of a sine wave can be represented with equation A.2, in which f_{carr} is the carrier frequency, f_{mod} the modulation frequency, A_c the carrier amplitude, A_m the amplitude of the information signal and Δf the frequency deviation. The frequency deviation is the maximum difference between the instantaneous frequency of the FM wave and the carrier frequency. If for example, the carrier frequency is 1600 Hz and the frequency deviation is 800 Hz, the instantaneous frequency of the FM signal will vary between 800 Hz and 2400 Hz.

$$y(t) = A_c \cdot sin\left(2\pi f_{carr} t + \frac{A_m \Delta f}{f_{mod}} sin(2\pi f_{mod} t)\right) \tag{A.2}$$

[10] Source of image: https://github.com/PetarV-/TikZ/tree/master/Amplitude%20modulation

B

Sound Quality Metrics Results

In this appendix the five sound quality metrics are given for each aircraft type. The amount of flyover measurements for each aircraft type ranges from 1 measurement to 72 measurements. For aircraft types that have more than one measurement, the average of the sound quality metric was computed and used in the analysis of chapter 5. The calculated results can be seen in table B.1.

Table B.1: Sound quality metrics (average) for each aircraft type

Aircraft	N_5 **[sone]**	K_5 **[t.u]**	R_5 **[asper]**	S_5 **[acum]**	FS_5 **[vacil]**
Boeing 737-300	104.8643	0.2879	0.1371	1.8251	0.4633
Boeing 737-400	108.5545	0.2667	0.1383	1.8025	0.6451
Boeing 737-500	95.4546	0.2549	0.1159	1.9019	0.4488
Boeing 737-700	104.9888	0.2381	0.1382	2.0221	0.4440
Boeing 737-800	117.9782	0.2325	0.1146	2.0075	0.4332
Boeing 737-900	118.4797	0.2272	0.1398	2.0167	0.4341
Boeing 747-400	158.1477	0.2798	0.1462	1.7246	0.5045
Boeing 747-8	156.7752	0.2872	0.2174	1.8289	0.9553
Boeing 767-300	113.8969	0.2536	0.1944	1.7045	0.4110
Boeing 777-200LR	137.8083	0.2775	0.1397	1.9687	0.4857
Boeing 777-300	143.1605	0.2886	0.1059	1.9665	0.3695
Boeing 787-800	102.6676	0.2547	0.0785	1.8680	0.3445
Boeing 787-900	116.2653	0.2407	0.0918	1.8835	0.3810
Airbus A300-600	121.2846	0.2494	0.1314	1.7662	0.8802
Airbus A319	103.6459	0.2931	0.0870	2.0295	0.3858
Airbus A320	101.5242	0.2935	0.0880	1.9993	0.3908
Airbus A321	107.7911	0.2680	0.8986	1.9612	0.4016
Airbus A330-200	130.9947	0.2416	0.0777	2.1152	0.3229
Airbus A380-800	127.9333	0.2419	0.1578	1.9024	0.5349
Embraer E145	73.8182	0.2611	0.1452	1.8689	0.5424
Embraer E175	96.9038	0.2883	0.0831	2.0902	0.4031
Embraer E190	98.7007	0.2428	0.0958	1.9813	04073
Fokker F70	79.2457	0.2598	0.1705	1.8108	0.4891
Bombardier CRJ900	86.6205	0.2692	0.0785	2.0849	0.3894
Bombardier CRJ1000	97.3173	0.2247	0.0798	2.1012	0.3036
Avro RJ-85	93.8809	0.2387	0.0851	2.4999	0.3400

C

CORRELATIONS OF DESIGN PARAMETERS WITH SQ METRICS

This appendix give the remaining correlations of the design parameter analysis. Five sound quality metrics were analyzed using five design variables, amounting to a total of 25 correlations. Four of them were presented in chapter 5 and the remaining 21 correlations are presented in this appendix.

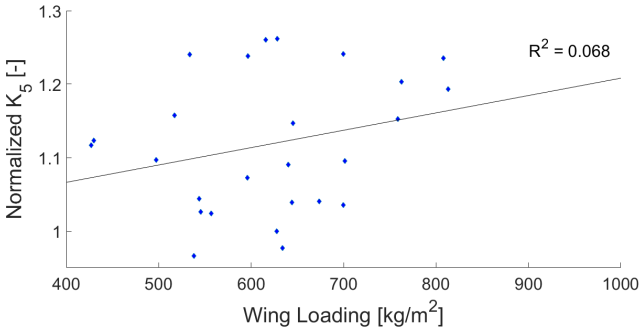


Figure C.1: Correlation of wing loading with tonality

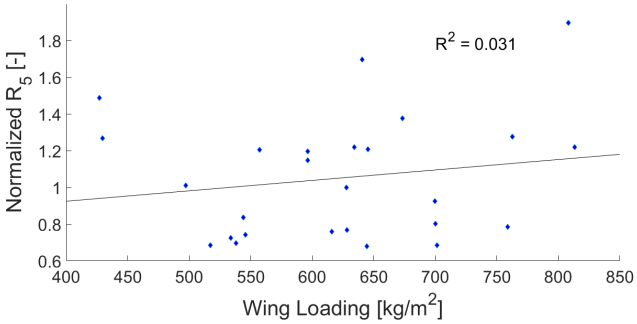


Figure C.2: Correlation of wing loading with roughness

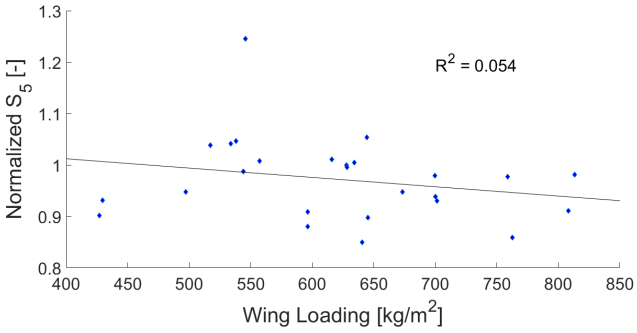


Figure C.3: Correlation of wing loading with sharpness

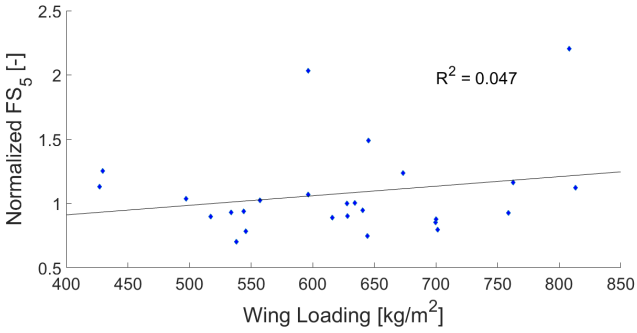


Figure C.4: Correlation of wing loading with fluctuation strength

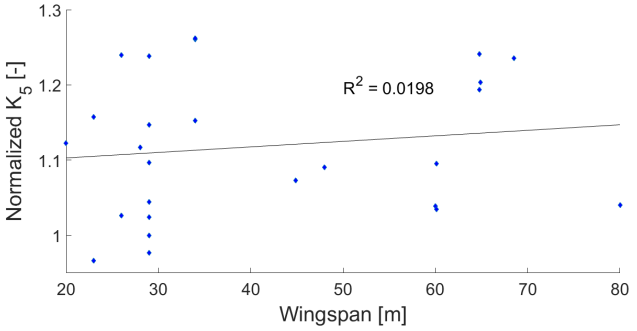


Figure C.5: Correlation of wingspan with tonality

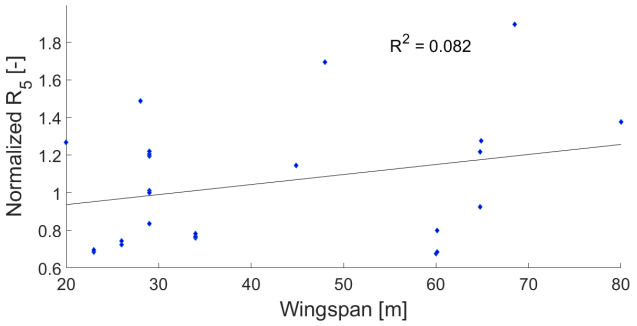


Figure C.6: Correlation of wingspan with roughness

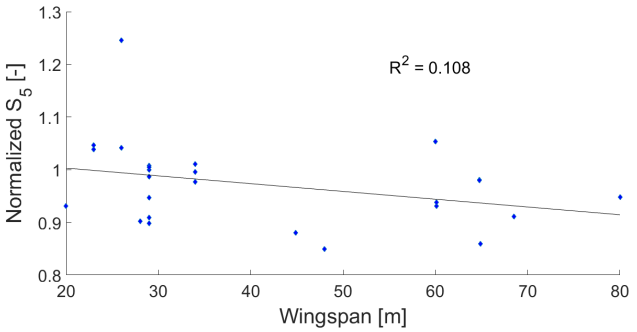


Figure C.7: Correlation of wingspan with sharpness

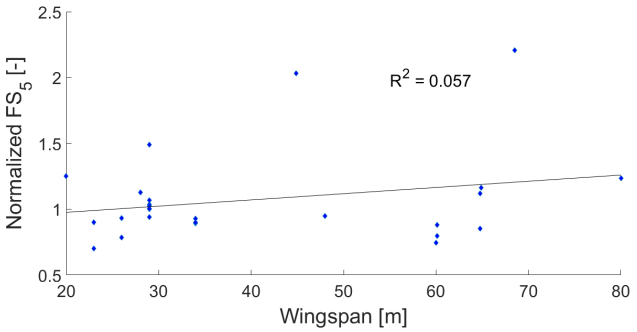


Figure C.8: Correlation of wingspan with fluctuation strength

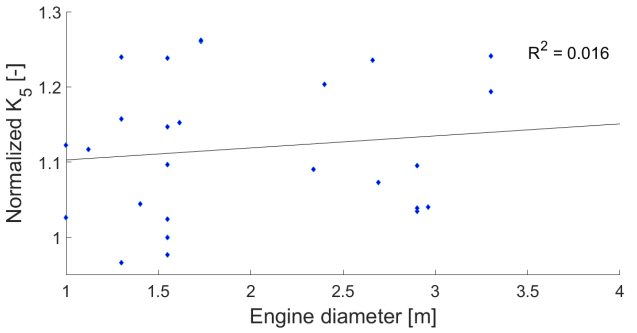


Figure C.9: Correlation of engine diameter with tonality

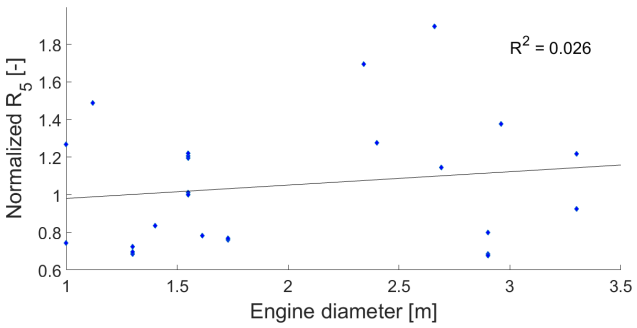


Figure C.10: Correlation of engine diameter with roughness

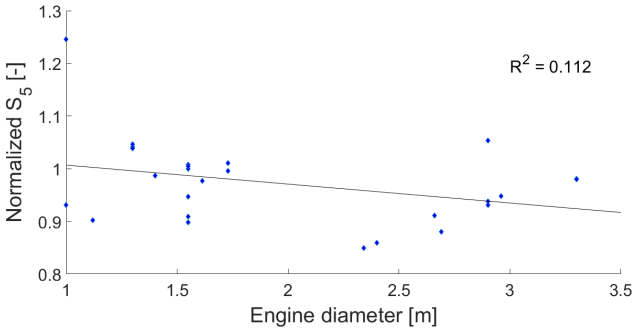


Figure C.11: Correlation of engine diameter with sharpness

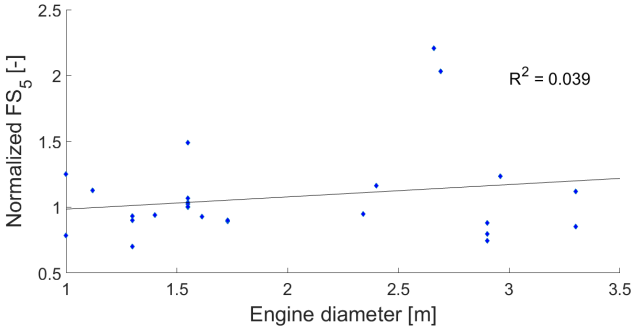


Figure C.12: Correlation of engine diameter with fluctuation strength

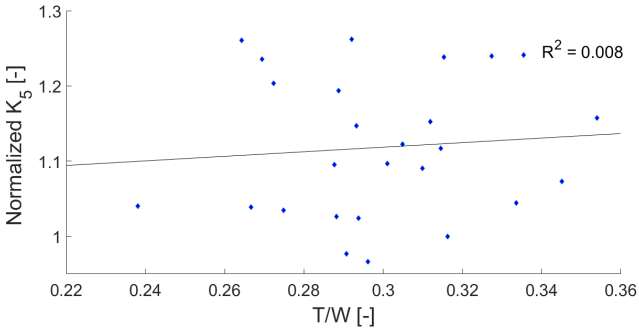


Figure C.13: Correlation of thrust-to-weight ratio with tonality

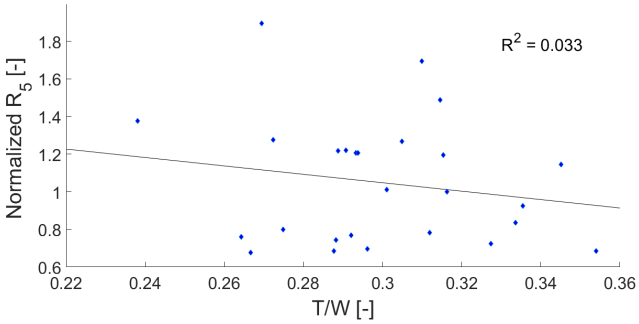


Figure C.14: Correlation of thrust-to-weight ratio with roughness

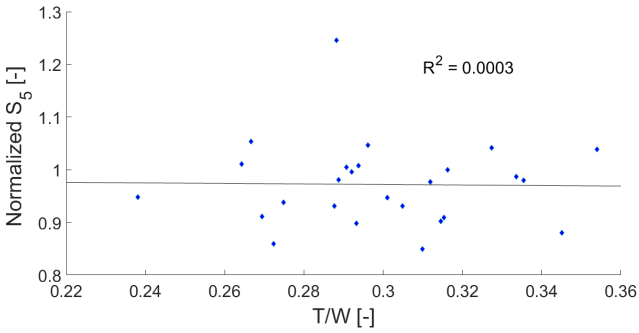


Figure C.15: Correlation of thrust-to-weight ratio with sharpness

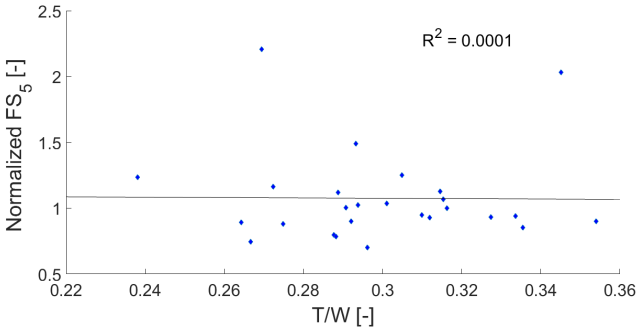


Figure C.16: Correlation of thrust-to-weight ratio with fluctuation strength

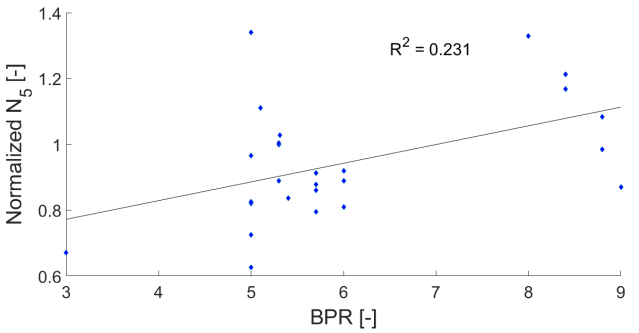


Figure C.17: Correlation of bypass ratio with loudness

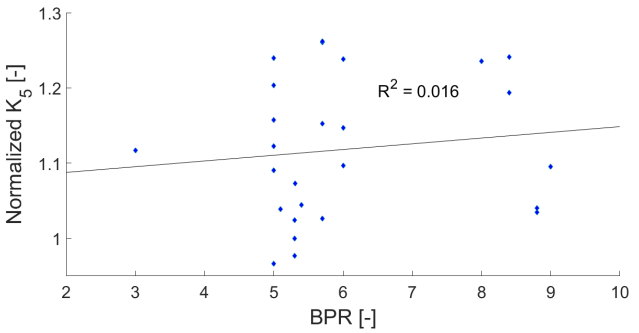


Figure C.18: Correlation of bypass ratio with tonality

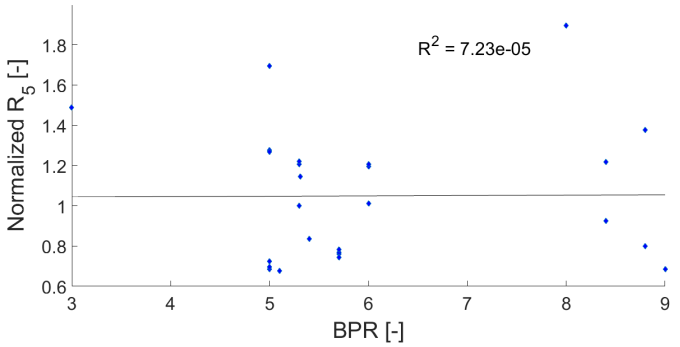


Figure C.19: Correlation of bypass ratio with roughness

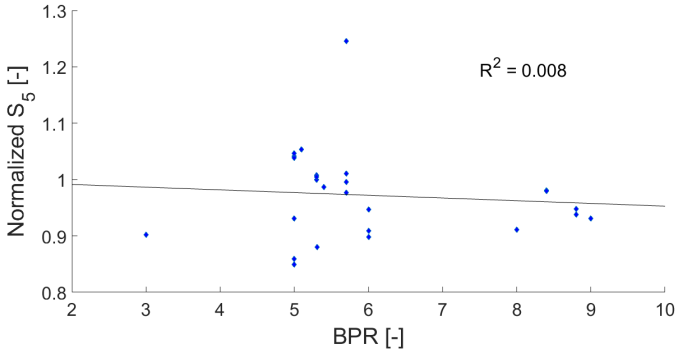


Figure C.20: Correlation of bypass ratio with sharpness

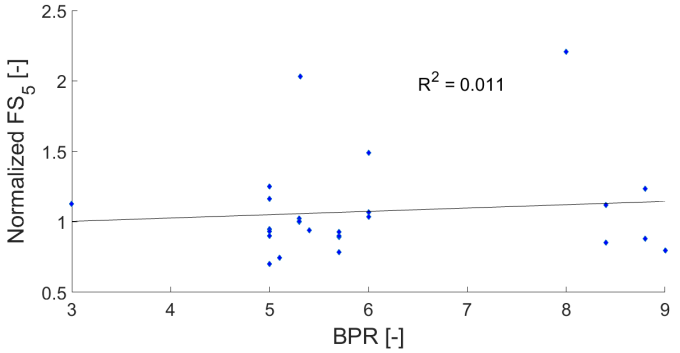


Figure C.21: Correlation of bypass ratio with fluctuation strength

D

DIRECT SCALING LISTENING TEST RATINGS

This appendix gives the complete results of the direct scaling test. Ten aircraft flyovers can be seen with their PA_{mod} and EPNL values. All the ratings given by the 20 subjects are also presented in the table. In the last column the calculated average rating can be seen. The average rating has been used in figures 5.12 and 5.13.

Table D.1: Results direct scaling test

Sound	PAmod	EPNL	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15	Subject 16	Subject 17	Subject 18	Subject 19	Subject 20	Average
1	167.22	101.31	75	30	20	50	50	50	75	40	10	25	60	30	30	35	50	30	40	40	25	40	40.25
2	205.38	105.36	85	50	25	60	75	75	85	40	50	40	80	50	26	65	65	35	50	60	50	60	56.3
3	235.25	106.43	85	43	30	60	75	75	86	60	60	60	80	50	35	70	66	45	55	65	50	60	60.5
4	116.84	96.90	50	20	20	40	75	25	70	30	20	60	40	60	15	20	45	25	40	45	10	10	36
5	203.23	103.95	80	65	50	70	80	100	78	50	50	65	70	60	30	70	75	60	70	70	30	75	64.9
6	189.24	103.24	65	55	50	69	80	50	85	65	60	65	60	40	40	70	75	55	75	65	45	74	62.15
7	236.51	107.42	85	70	51	75	80	100	85	75	80	65	80	40	29	75	85	75	80	80	55	70	71.75
8	318.02	115.63	100	90	75	76	100	100	90	85	90	90	90	80	47	90	100	85	90	99	75	85	86.85
9	203.74	104.65	75	75	70	70	100	75	87	60	70	65	70	50	36	80	90	60	90	82	15	80	70
10	262.06	107.99	80	80	75	70	75	75	88	60	80	75	75	40	35	85	85	80	95	92	40	75	73