Towards Innovative Ways to Assess Annoyance

**Catherine Lavandier , Roalt Aalmoes , Romain Dedieu ,
Ferenc Marki , Stephan Großarth, Dirk Schreckenberg ,
Asma Gharbi , and Dimitris Kotzinos**

Abstract Technological changes have driven the developments in the field of noise annoyance research. It helped to increase knowledge on the topic substantially. It also provides opportunities to conduct novel research. The introduction of the internet, the mobile phone, and miniaturisation and improved sensor technology are at the core of the three research examples presented in this chapter. The first example is the use of a Virtual Reality simulation to evaluate aircraft flyovers in different environments, and it examines how visual perception influences noise annoyance. The second example describes the use of a mobile application applying an Experience Sampling Method to assess noise annoyance for a group of people living near an airport. The third and final example is a study over social media discussions in relation to noise annoyance and quality of life around airports. These three examples demonstrate how novel

C. Lavandier (✉) · R. Dedieu · A. Gharbi · D. Kotzinos
CY Cergy Paris University, Cergy-Pontoise, France
e-mail: catherine.lavandier@cyu.fr

R. Dedieu
e-mail: romain.dedieu1@cyu.fr

A. Gharbi
e-mail: asma.gharbi@ensea.fr

D. Kotzinos
e-mail: dimitrios.kotzinos@cyu.fr

R. Aalmoes
Royal Netherlands Aerospace Centre NLR, Amsterdam, The Netherlands
e-mail: roalt.aalmoes@nlr.nl

F. Marki
Budapest University of Technology and Economics, Budapest, Hungary
e-mail: marki@hit.bme.hu

S. Großarth · D. Schreckenberg
ZEUS GmbH, Hagen, Germany
e-mail: grossarth@zeusgmbh.de

D. Schreckenberg
e-mail: schreckenberg@zeusgmbh.de

technologies help to collect and analyse data from people who live around airports, and so improve our understanding of the effect of noise on humans.

Introduction

Noise annoyance is a known health-related societal problem for a long time, and cannot be solved in the near future. But the conditions and environment in which it occurs do change. Significant changes have taken place in the twentieth century that affect how noise annoyance occurs and how it is perceived, and this trend continued in the twenty-first century. If we look at the forces that come into play, we can discriminate four interrelated factors (Fig. 11.1).

First, the noise sources change: if we look at the transportation domain, new vehicles appear (such as drones), or existing vehicles go through a disruptive cycle, such as the movement from petrol-based towards electric-powered automobiles. Research on noise mitigation measures has also reduced the impact by noise at the source. For instance, by increasing the by-pass ratio of jet engines, a significant noise reduction is achieved, making aircraft much more quiet than earlier generations (see Chap. 5 and Fig. 11.2).

Second, human perception and attitude, and consequently how people react to noise, have changed. Noise nuisance that was previously accepted and considered part of the environment is, with a more vocal community, noticed much more and may lead to complaints that are more significant. In one way, the improved democratic instruments or government protection measures enhance the ability to complain about noise issues, but new and more efficient ways of communication cannot be ignored

Fig. 11.1 Noise sources that cause annoyance have changed during the course of time. Human perception and the attitude towards noise annoyance have changed as well. Technological developments and expanded knowledge have both driven these changes on both noise sources and human perception and attitude
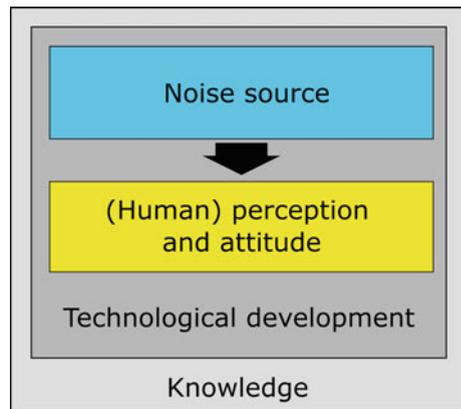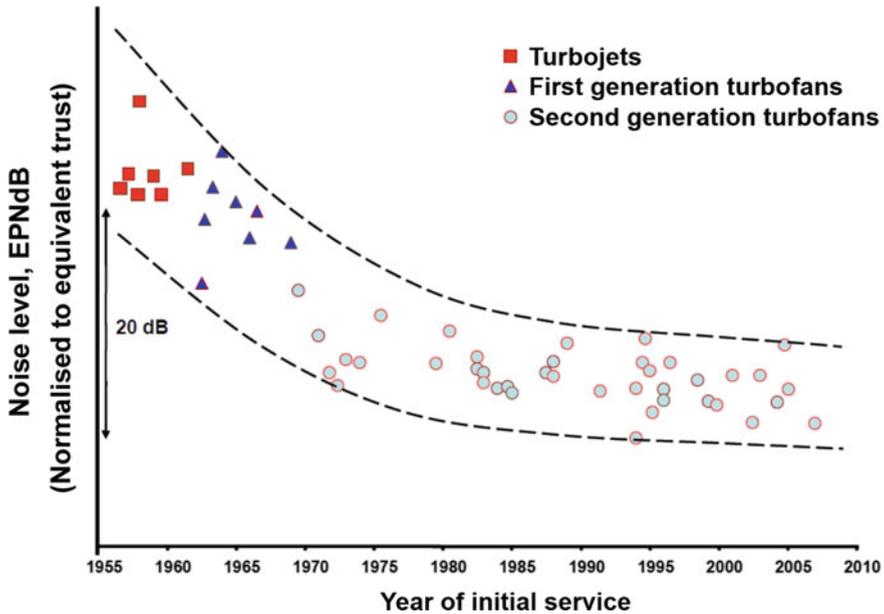
Fig. 11.2 Trends of noise reduction through the twentieth century. *Source* Le livre blanc de l'acoustique en France en 2010 (edited by SFA, the French Acoustical Society)

as well. With the internet, people are more organised to complain via email, websites, and social media, and authorities are also more organised to communicate with them and enhance people's engagement.

Indeed, aircraft noise annoyance per given intensity of average sound levels have increased over the last decades although single airplanes got quieter at the same time [14]. As described in Chapt. 9, there is evidence that annoyance mediates the impact of noise exposure on further long-term health risks, indicating that the increase in annoyance over time would in the long run also affect long-term health effects of aircraft noise.

The third reason is the gain in knowledge on noise annoyance on humans. The vast amount of research undertaken improved both psycho-acoustic knowledge and the impact on health. It created standardised exposure–response curves and standardised noise annoyance research questionnaires, such as the ICBEN scales [10]. And, the improvements in noise effect research led to human health reports on noise impact such as that by the WHO [46].

The fourth reason for the change of environment is the change of technology: technological advances reduced noise at the source, made people more aware of noise nuisance, and helped increase knowledge on the topic of noise annoyance. But it will also provide researchers with new means to conduct noise research. The introduction of the internet with email, websites, and, subsequently, social media creates a larger audience for conducting large-scale evaluations. The introduction and

large-scale adoption of (smart) cell phones creates a potential pool of test subjects that can register their location, record sounds, and answer questions related to local soundscapes and noise annoyance. Finally, miniaturisation, improved processing power, and development of sensor technology has led to revolutionary technologies such as Virtual Reality and Augmented Reality. Typical examples of these devices are the Google Glasses, the Oculus Rift, and the Microsoft Hololens. In addition, research projects may use advanced virtual-reality headsets based on the combined use of head-tracking and Head Relay Transfer Functions (HRTFs) to provide natural spatial audio representation. The combination of visual and audio stimuli can create an immersive simulation to mimic real-life experiences that could otherwise only be examined using empirical studies.

In the next sections, we focus on some innovative ways to conduct acoustic research towards noise annoyance. In no way are these examples exhaustive, in the sense that they cover all recent technological innovations; but hopefully, they would help in showing how they can help finding new ways to characterise, mitigate, or manage annoyance; Or conducting other new innovative research in the domain of acoustics or in related fields. The first example is the use of a Virtual Reality simulation to evaluate aircraft flyovers in different environments, and it examines how visual perception influences noise annoyance. The second example describes the use of a mobile application to apply an Experience Sampling Method to assess noise annoyance for a group of people living near an airport. The third and final example is a study over social media discussions in relation to noise annoyance and quality of life around airports. The last two examples could also be combined with the dynamic population maps that are described in the following chapter, to correlate people's location and their annoyance, a novel approach not seen previously in aircraft annoyance research.

Immersive Simulation to Mimic Real-Life Experiences

Communication and engagement by airport authorities, local government, or local planners with communities is important and is discussed in other sections of this book. With respect to communication on noise impact, a more difficult task is at hand to explain predicted noise levels and what they mean for the affected communities. There are different ways to present changes to the noise on paper, and those used often are the 24h annual noise level, single-event peak-levels, or number of (highly) annoyed people near the airport. But to make these numbers better comprehensible for the layman, a demonstration that simulates aircraft flyovers at the predicted sound level would clarify what these numbers mean. Novel approaches that use aural and visual stimuli can be used for this purpose. There are some virtual reality applications (auralisation and visualisation) which can be used by residents to give them better understanding of the impact of future airport scenarios in land-use planning, as the virtual reality creates a higher immersion for the user. Virtual Reality headsets feature a greater field of view than projector or TV screen. Additionally, using head-tracking

sensors, they allow the system to change the visuals in a way that allows the user to look in all directions. The same is true for the (spatial) sounds that are produced and provide audio-directivity. But this immersion has to be ecologically valid if authorities want to be trusted by residents during communication campaigns.

Validation of a Virtual Reality Application for Aircraft Noise

In this section, a validation study [8] is presented to evaluate the Virtual Community Noise Simulator (VCNS) for the perception of aircraft noise (Fig. 11.3). The VCNS has been developed by the Netherlands Aerospace Centre (NLR) based upon earlier work done at NASA Langley [38]. The current set-up makes use of the shelf hardware, such as an Oculus Rift CV1 VR headset, supported by a powerful laptop computer, and separate headphones for the audio.

Participants of the study take part in a perceptual experiment. Two landscapes are presented in which the participants evaluate the flyover sounds and visuals of three distinct aircraft. These three aircraft are the Airbus A320neo, the Airbus A380, and a revolutionary design called the "BOLT" (Blended wing body with Optimised Low-noise Technologies, Fig. 11.4, see also Chap. 6). The influence of the visuals is also measured by presenting the sound of one aircraft with the visuals of the other aircraft as well. In one additional condition, the visual is not visible and this is represented by an overcast sky. To prevent influence from different background noises, a single background recording was used in both landscapes.

In order to test if the size of the aircraft (or the absence of the aircraft vision because of clouds) has an influence on the perception of the audio-visual situation, all synthesised sounds were crossed with all visual situations. So, twenty-four environmental situations were created (three types of aircraft sound x four types of visual

Fig. 11.3 Artist impression of the virtual community noise simulator by NLR
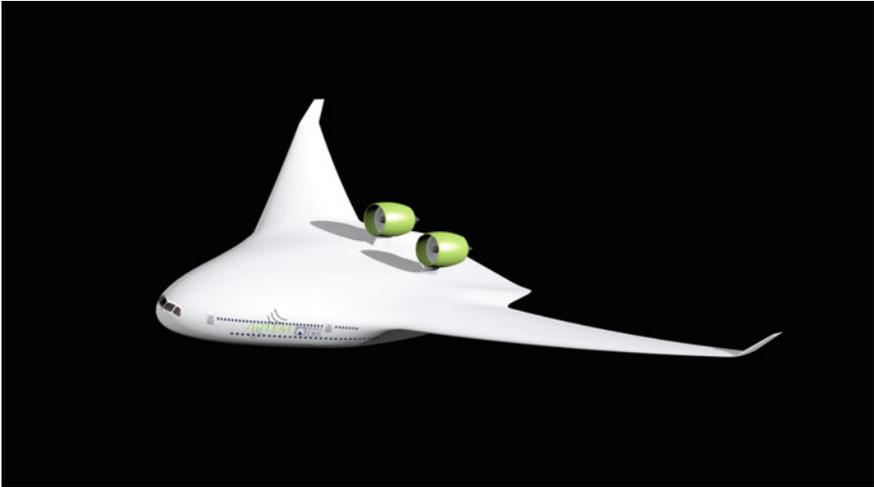
Fig. 11.4 The "BOLT" architecture

aircraft source x two types of landscape). The four types of visual aircraft source correspond to the three aircraft plus one situation with clouds in the sky. The two landscapes correspond to one green park, and one urban situation.

Sixty participants were immersed in these twenty-four situations. After each flyover, they had to use the joystick on the touch controller to give answers to a questionnaire which appeared in the virtual world (Fig. 11.5). It consisted of four ratings:

(1)   Overall, does this situation seem more or less

Unpleasant/Unbearable ….. Pleasant/ Bearable?

(2)   Does the association of sound with visual seem more or less

Unrealistic/Non credible/Incoherent ….. Realistic/Credible/Coherent?

(3)   Is the sound of this aircraft more or less

Unpleasant/Unbearable ….. Pleasant/Bearable?

Fig. 11.5 Question about the audio-visual situation in the virtual world

(4)   Does the noise level of this aircraft seem more or less

Strong/Loud ….. Weak/Quiet?

After assessing the twenty-four situations, participants had to fill a final questionnaire concerning the simulator's overall efficiency in creating a feeling of reality, and concerning personal information such as their noise sensitivity, or their quality of life. As the experiment was conducted by the Cergy Paris University (CYU, France), it was previously submitted to this ethical committee which approved it.

For the overall pleasantness measure, three groups of participants rated the audio-visual situations differently. One group rated all the situations in the "negative" part of the scale, which means that they found the environment less pleasant than the other groups. Generally, participants of this group are more noise sensitive than the other participants and a little bit older. Another group of less noise sensitive people rated the overall pleasantness in the middle of the scale. The last group rated the overall pleasantness in the "positive" part of the scale. In this group, participants are a little bit younger, whatever their noise sensitivity.

If we have a look on the influence of the landscape, it seems that the majority of participants were not influenced by the landscape, but some of them preferred the flyover in the park, because the situation is greener [45] and some of them disliked the flyover in the park because the flyover disturb the quietness of this environmental situation [4]. The size of the aircraft has no influence on the sound perception nor on the overall pleasantness.

For the sound pleasantness, the aircraft sound of the A380 is the most unpleasant because it is the loudest one ($L_{Amax,30 s}$ = 76.1 dB(A)). This sound also has the lowest pitch. Then the A320neo and the new aircraft are perceived as more pleasant as they are both less noisy ($L_{Amax,30 s}$ = 72.1 dB(A) and $L_{Amax,30 s}$ = 71.3 dB(A) respectively). People react globally in agreement with results of the literature about sound perception [12, 26, 30]. The study about realism can explain why the results are so close to scientific literature about aircraft noise.

In total, 78% of the participants found the virtual audio-visual environment very or extremely realistic (Fig. 11.6). All sound syntheses are of similar real-
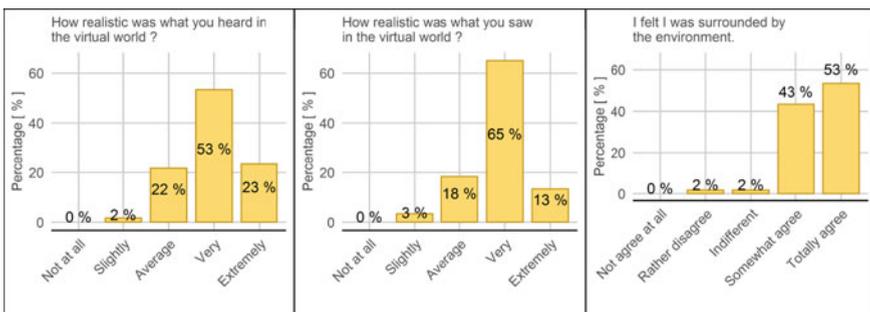

Fig. 11.6 Distribution of the answers given by participants at the end of the experiment about realism and immersion

istic/credible/coherent quality. Nevertheless, participants noticed that the simulation of the A320neo is the most credible from a visual point of view. The odd architecture of the "BOLT" aircraft renders its visual simulation less credible, the A380 is also less credible than the A320neo because people are less familiar with this large aircraft. It has also been found that the realism of the visuals are not only deteriorated by the unknown nature of the aircraft but also by clouds. Moreover, presence of clouds leads to an overall more unpleasant situation, and the sound of the new aircraft is judged louder under clouds compared to the same sound coming from any aircraft that is seen in the sky.

To conclude, the quality of the application has been validated by our perceptual experiment. 96% of the participants felt surrounded by the environment and the results, which are in line with literature about aircraft noise, show that this virtual reality tool can be used for communication with residents around airports in a fair approach. If we want to improve the quality of the application, the efforts should focus on the visualisation of the artificial clouds.

Effectiveness of This Virtual Reality Application for Better Communication

In order to test the effectiveness of this simulation tool, an in-situ experiment will be organised in a city around an airport where an operational change of the aircraft route is planned. If people feel that they understand this change better with the virtual reality tool than with classical maps (noise maps and aircraft trajectories), in theory they should feel in more control and thus should be more confident with the airport authorities. The hypothesis behind this in-situ experiment is that the noise annoyance could be reduced with the use of such a tool, reducing fear about what will happen in the future.

Mobile Application to Assess People's "Annoyance"

Introduction

For about two decades, airport residents' long-term annoyance has been studied by asking the ICBEN question, "Thinking about the last 12 months or so, how much did aircraft noise as a whole bother, disturb or annoy you?" [10]. Typically, and in line with the ICBEN recommendations, annoyance ratings have been given on 5-point verbal scales and 11-point numerical scales. By dichotomising the answers in values of high (1) and not high (0) annoyance, the percentage of respondents highly annoyed related to computed average noise levels provide exposure–response curves
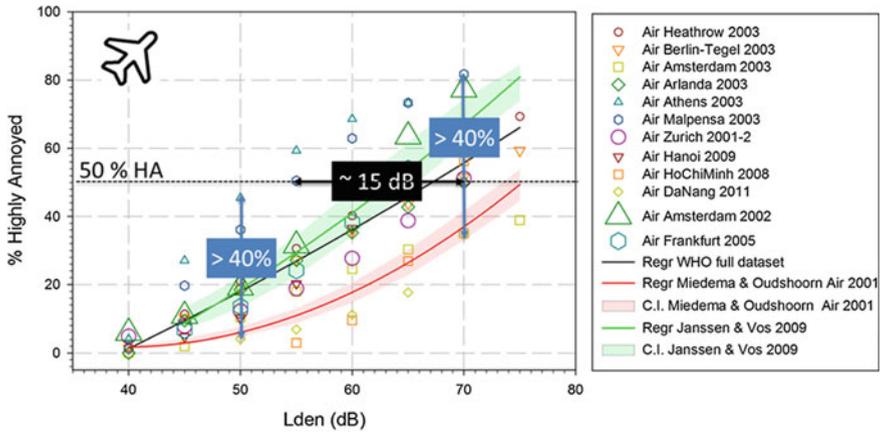
Fig. 11.7 Variance in exposure–response curves for the percentage of persons highly annoyed by aircraft noise [15]. *Source* Guski et al. [15], adapted and modified

that inform noise policy. Unfortunately, various study exposure–response curves deviate significantly (Fig. 11.7).

During these field studies, researchers also asked participants other questions to find reasons for their specific level of annoyance. However, here again, significant differences occurred. While there have been found many important driving factors to get annoyed, after decades of aircraft noise annoyance research, we are still not in the position to be able to set mathematical models, which would estimate with acceptable accuracy the annoyance due to its contextual embedment. It turned out that non-acoustic factors play an important role in the judgements [13], but the weight of one or the other factor is hugely different from place to place, and situation to situation, because airport residents and their living conditions are different. Those who live near a huge hub, like Heathrow or Frankfurt, behave differently from those who live near medium or small sized airports. Whether the airport has night traffic or not is also important. Over the years it has also been found that people in a changing situation (e.g. a new runway or a new flight route) behave completely differently as opposed to those living around a steady state airport. All this leads to the conclusion that while there have been quite a lot of field studies conducted already, because of the diversity of situations, we still don't have enough information to predict annoyance.

However, these classical field studies are quite costly and furthermore it has increasingly become difficult to get participants to take part. So, we need new ways to make test procedure and annoyance estimation easier, thus less costly and to get "access" to a lot of people, to find the necessary number of volunteers to take part. And therefore mobile phones come into play [24, 34, 37]: through them, people are very easily be accessed with regard to their short-term annoyance, the test procedure can be much more flexible (e.g. asking participants not just once but at several random times is easily done by an appropriate mobile application) and data exchange between

researcher and participant is also fast and reliable. Moreover, acoustic measurements can be stored on the device, at the assessment moments.

It seems that it is worth trying to modify also the methodology, because the importance of non-acoustic factors means, at the end, that the key problem is the deterioration of the quality of life of people. Further, annoyance can be understood as a stress response to noise (see Chaps. 9 and 10) driven by the noise exposure as well as the perceived lack of control or capacity to cope with the noise. In addition, when being asked retrospectively about the noise annoyance over a period of several months as recommended by ICBEN, annoyance judgments could be biased by memory capacity (see also Chap. 10). Therefore, a momentary assessment, which captures feelings in real time, became the choice for ANIMA's pilot study to assess sound perception in airport regions closer in time to the sound events.

The ANIMA Mobile Application (AnimApp)

Experience Sampling Method

The method proposed—in order to assess acute perception of sound events—is called Experience Sampling Method (ESM; sometimes also called Ecological Momentary Assessment). In contrast to the assessment of retrospective long-term judgments as it is done in most of the 'classical' socio-acoustic surveys, this method allows to assess the experiences in-situ repeatedly on different (consecutive) days at different times of day. Hence, the ESM approach can be characterised as "capturing life as it's lived" [3]. By installing a survey-software on participants' devices, researchers are able to prompt for several assessments, whenever it appears to be necessary. Data is then submitted to a server and is available as soon as the upload is finished. Although ESMs have been found to be useful in many scientific disciplines they have just shortly found their way into modern noise impact assessments [7, 11, 28]. Here, we have found a promising way to get a realistic insight into people's everyday noise experience, which we regard as essential when examining sound perception and its impact on quality of life in individuals.

The ANIMA project features among others the development of a new mobile application, called ANIMA Research, nicknamed simply AnimApp (see Fig. 11.8). Using the application, we carry out an ESM study about the impact of the sound-scape and landscape of the surrounding environment on people's perception of the environment and their quality of life around airports.
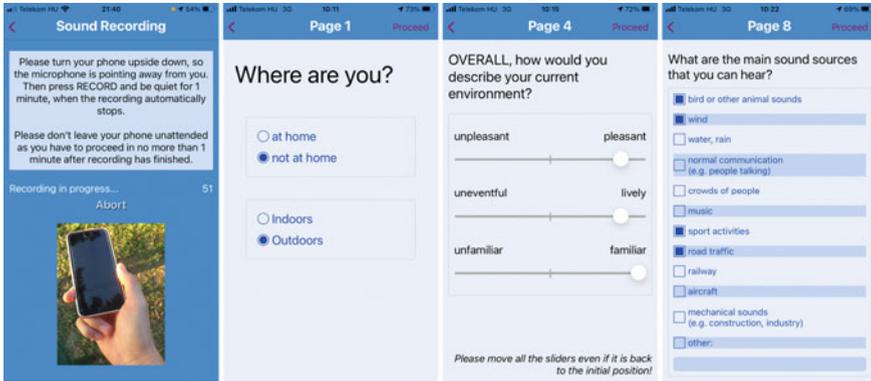
Fig. 11.8 Example screenshots of the Anima Research application 'AnimApp'

Global Structure of the Application

Explanations and Permissions

After installing the app, during the first start, the test procedure is explained, permissions are asked for location and microphone usage, and notification sending. In addition, settings have to be reviewed and adjusted at will, so the frequency and time-span of the test suits the participant's lifestyle. Then the user exits the app. The application is designed as self-operating. Once installed, notifications will prompt for assessments at random intervals. The selection of momentary, weekly or the final questionnaire as well as the replacement of missed notifications are all automatically done without further intervention by the participants or the research team.

Momentary Assessments

During weekdays, from 7 A.M. until 11 P.M. (when not shortened by the participant), once around each full hour the app sends a notification (Fig. 11.9):

Depending on the user's preference 2 to 4 assessment notifications come a day. Hours of measurement (i.e. 7 A.M, 8 A.M, etc.) are randomised for the whole duration of the test, so the user doesn't know when the next measurement request will be prompted for. Each hour of the day is tested within the adjusted interval. The total duration of the study adapts correspondingly.

The user has to respond to a notification in 20 min and to start the momentary assessment consisting of sound recording and questionnaire filling.

During weekend days, the participants respond to the same questionnaire, however in a shorter time frame, i.e. from 10 A.M. till 10 P.M. and just every second hour only.
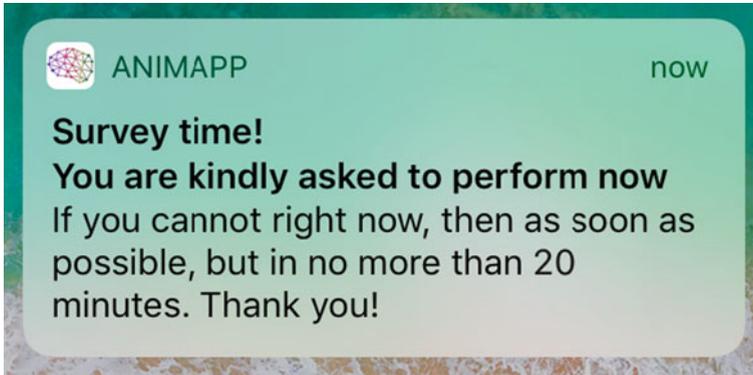
Fig. 11.9 Notification calling to perform an assessment

End-Of-Week Assessments

At the end of the work week, i.e. on Friday evening, a short end-of-week questionnaire has to be filled in. In addition, after the very last weekend's momentary assessment, the same end-of-week questionnaire is asked relating to all *weekend* days during the test.

Final Questionnaire

Once all week and week-end hours have been performed, a final questionnaire is presented to the user asking for noise sensitivity, and standardised questions on well-being [17, 33].

Acoustic Measurements and Selected Questions

The ANIMA project tries to depart from the classical approach by moving from the focus on average noise pollution and annoyance towards a broader view and the general notion of the perception of acute sound quality and its impact on quality of life. The broadened, more open content of judgments is combined with its assessment in acute, specific moments making the response more independent from memory bias and from biases that may come along with the noise-attributing wording (so-called demand characteristics) of the standardised long-term annoyance questions.

Regarding the acoustic metric describing the sound environment, it is worth trying to find better acoustic metrics, which are closer related to people's sound perception than the day-evening-night sound level L_{den}, which summarises and weights noise events over a 24 h period of the day. In order to be able to calculate most indicators

that are proposed in the literature, the spectrum of the recorded sound (third octave band, each second) is stored in the AnimApp.

At each moment when a notification is received by a participant, a 1-min acoustic measurement has to be performed and then a series of questions appears. The questions that have been selected are inspired from the so-called *soundscape* questionnaires [2, 5], which capture all relevant dimensions that can explain the impact of sound environment on people and which have been recently standardised [18, 19]. The first dimension is the pleasantness of the sound, followed by the eventfulness, and the familiarity with the environment. The acoustic environment should also be described with the types of sound sources, which are present in the environment. The context is not limited to the location (which is captured by the smartphone), but should also concern the activity of the participant at the moment of the evaluation. The context also has to include visual data. Actually, it has been shown that the quality of the landscape has an influence on the perceived pleasantness of a soundscape when people are outside: the greener the landscape, the higher quality the soundscape [31, 32, 45]. When individuals are inside, the natural elements people could see through their windows reduced the negative effect due to noise [43]. We examine this further by asking participants to answer what they see through their windows, if they have a view of the outside. In the frame of our approach, we also want to question the rating of long-term annoyance by means of single items: people feel disturbed at different moments of the day, or evening, or even night, but participants could have difficulties to produce a valid annoyance rating over a longer period time (e.g. 12 months) [9, 40, 41, 44]. To examine how people add up all the different experiences deriving from their perception - at least for a one week period, we decided to ask 3 questions on the environment (overall impression on the sound pleasantness, landscape pleasantness, and representativeness of the week) at the end of each week.

Of course, the unexplained variance of noise annoyance could partly derive from personal dispositions. Accordingly, we assess the mood at each notification. Furthermore, the individual noise sensitivity and the perceived quality of life is assessed in the final questionnaire.

Development of the Mobile Application

AnimApp was developed for the two operating systems Android and iOS. This allows a widespread use of the application on the vast majority of modern smartphones in use [39] keeping participation requirements on low threshold. For the use of AnimApp on these two operating systems, several operation system specific adaptations and adjustments were applied.

In regular field studies, the procedure of the study is explained in detail to the participants and once they agree to participate, they tend to comply well, which can be enforced by offering an expense allowance. But with mobile applications, long text-based explanations, which might demotivate participants to continue participating have to be avoided. Specific effort has to be deployed in formulating instructions precisely and shortly at the same time.

Another difficulty is how to bring our test through, when the user often disregards the notifications to do an assessment. After a pre-test phase, we decided to allow 20 min delay between notifications and possible answers, and to send reminder notifications each 5 min to the user.

The regularity of data provision also needs attention: on the one hand, apps should avoid running all the time in the background (and thus draining battery), but on the other hand they must make sure to send assessment data to us, once a measurement is completed. In our case, when connectivity is not available after fulfilling a measurement, there is no other option than to schedule data sending for later time. However, we ensured the battery will be drained as little as possible and even if the application is killed by the user the schedule for the upload of data still persists.

Finally, tracking of participants' location imposes further potential problems. In our study we want to know the position of the user at the time of the assessment, so that we can estimate the aircraft noise exposure for the respective positions afterwards. Additionally, we ask our participants to allow tracking of their position all the time, so we have an impression how airport residents move during the day (i.e. to know—based on noise maps—how much they are exposed to aircraft versus other noise). This option needs consent of the user (see below; paragraph on data security and privacy).

Randomisation

For all field studies, from the point of view of later statistical analysis, the randomness of sample collection is very important. Therefore, to assure good randomisation among assessment hours and among participants, for AnimApp it has been decided (a) to let the participants perform an assessment at randomly selected hours (but along the test, each hour will be assessed just once), each day 2–4 times depending what he/she set up in the settings, (b) to define a 10 min time-frame around full hours and then randomly select the exact time in the resulting time span (e.g. between 7:55 and 8:05).

Data Security, Privacy

AnimApp made several steps to respect people's privacy and to be fully compliant to GDPR:

- It is not necessary for the users to enter any personal data to register in the study, they simply get automatically the next free user ID, thus users remain anonymous, and their answers too.
- The sound recording is right on the phone transferred into a series of 3rd octave band spectra, one for each second, and only this is transmitted to the server. This keeps privacy as the original audio recording cannot be reconstructed from these acoustic data.

- The user has explicitly to agree to constant location tracking, and can also refuse if preferred. Also, positions are rounded to a grid of 100 * 100 m on the user's phone and only then sent to the server.
- For any cases, during first use, a user must explicitly agree to our privacy policy, including the agreement that we collect/store/process data from the participant with his/her consent.

Experiences Around Two Airports

A first version of the application has been tested during winter 2018, and a feedback questionnaire has then been proposed to the "beta" testers. Based on these feedbacks, the test procedure has been refined and the final version (which has been described in this Chapter) will be used for the actual study. Two different sized airports will be observed. The application will be experimented during spring/summer 2021 around Ljubljana Airport in Slovenia, and London Heathrow Airport in the UK. Of course, the application has been translated into Slovenian for being used in Slovenia. Generally, instructions and indications are technically provided through separated libraries that are easing the adaptation to a wide range of other languages. Results should show whether such an approach could be used for more airports, and more suitable periods (more traffic for tourism, outside of a sanitary crisis like COVID-19).

Using Twitter as a Survey Tool: Understanding people's Opinions of Quality of Life Around Airports

Context

Social media has increasingly become a space where people meet to discuss, express opinions and debate over a wide range of subjects ranging from global politics to everyday life and from political and ideological opinions to advertisement of products and services. A specific part of the discussions about everyday life is the focus of this work, done as part of the ANIMA project but having wider applicability: this part concerns the understanding and subsequent classification of people's annoyance when they live, work or socialise around airports. In order to do this, we need to analyse discussions over an extended period of time which are somewhat localised since we need the involved people to either live or work around airports or to show a significant presence that would allow them to be considered as directly affected by the generated impact from the airport operation.

Actually, surveys assessing the impact of the operation of an airport on the population living in the adjacent areas have been carried out for a long time and have received particular focus (and a lot of scrutiny) in cases of airports' expansions.

These surveys have taken place in traditional ways, mainly through the selection of a representative part of the population which is contacted either by person (door-to-door), by phone or through post in order to fill a predetermined questionnaire. More recently, such surveys are being contacted either online or through mobile apps, where the invited subjects download and install an app on their mobile phones and then use the app to answer specific questions and/or allow it to monitor (parts of) their everyday life. Main disadvantages in both cases are the difficulty to extract big enough samples and to guarantee the participation of the users during the duration of the survey, since quality of life issues cannot be assessed in one-off answering. Moreover, as in all surveys, the usage of mobile apps raises various privacy concerns, which can, of course, be mitigated by extra developer effort, as it is the case in the ANIMA mobile app.

Compared to these methods, surveys based on social media research and analysis exhibit various advantages and disadvantages. On the advantages side, we can put the infrastructure for capturing social media posts once and then monitor the discussions for an extended period of time with no extra cost (besides the cost of processing and storage of the posts). Additionally, social media platforms like Twitter or Facebook can provide easy access to thousands or millions of users and millions or billions of relevant tweets (depending on the airport, the area, etc.) so as to extend the sample that "participates" in the survey. The users are actually taking part in online discussions based on their own interest, with no strict requirements. On the disadvantage side, online social media brings its own biases, for example it is well known that people from older generations use them very little or only for purposes of communication with family and friends. One more problem is that posts do not necessarily carry location information, so sometimes localising a discussion is not possible or becomes a costly operation by itself. Finally, discussions on social media are directly affected by whatever captures the public's eye as well as from the actual reality, for example during the recent COVID-19 crisis discussions on social media are overwhelmingly dominated by this and the lack of actual flights mitigated the issues and the discussions. The richness of information in social media can also be a curse: not all discussions are relevant to the specific subject. In that respect, we need first to extract the relevant posts or discussions, which is not a trivial subject by itself. Additionally, in the case of Twitter and other microblogging services the imposed limit on the number of characters for each post forces people to express themselves in unique and sometimes difficult to understand ways. Nevertheless, the number of posts that can be captured and the number of users that participate make it a viable alternative that—with the necessary scientific precautions—can provide valuable insights on the opinions and sentiments over quality-of-life issues around airports.

As part of the ANIMA project, we develop a set of reusable tools and methodologies that allow capturing the necessary relevant social media posts, extracting the topics of discussions around quality of life and classifying the sentiments around these topics as positive, neutral or negative trying to depict a qualitative assessment of the opinions of people from the area around airports. For the purposes of the project, we focus on the area around Heathrow airport in London, UK (one of the

busiest airports in a very densely populated area) but the methodology and principles described can be easily applied in any other similar case.

Scientific Background

The core of the work in this task is the extraction of opinions and the analysis of sentiments contained within those opinions that would allow us to classify those opinions as positive, neutral and negative.

Sentiment classification is a hard challenge that faces several challenges such as dealing with trivial posts, incomplete sentences, misspelling and abbreviation due to size restrictions, dealing with specific meanings such as irony or humour and the use of emotional expressions. Sentiment classification approaches can be classified into three main categories: (i) machine learning, (ii) lexicon based [22, 23, 47] and (iii) hybrid approach.

Machine learning-based sentiment analysis consists in predicting the polarity of sentiments by training a machine learning model with examples of emotions in text to automatically learn how to detect sentiment without human input. In the literature, one can find works that use emoticons [35], slang language and acronyms [16], words in text and their respective part-of-speech (POS). Other elements to consider are intensifiers such as all caps and characters' repetitions (e.g., happpyyy) [21], punctuation marks, n-grams [21] and negation marks [29] or (all possible) combinations [21] as features of the analysed tweets. In [35] and [6], authors performed a 2-way classification (i.e., positive or negative) on data with emoticons and applied respectively SVM (Support Vector Machine) and NB (Naive Bayes) algorithms that were able to achieve more than 70% accuracy.

Recent works tried neural networks with word embeddings for the representation of tweets and showed that they achieved much better performance in sentiment analysis [20, 27, 36, 42]. Word embeddings represent words by dense vectors with much lower dimensionality. Each word is positioned via its vector value into a multi-dimensional space (embedding space) which helps to consider their semantics (i.e., synonyms are geometrically close, antonyms are far from each other). Mathematical operations can also be applied on vectors and produces semantically correct results, e.g., the sum of the word embeddings of *king* and *female* produces the word embedding of *queen*. Ren et al. [36] has used a context-based convolutional neural network (CNN) to apply sentiment classification on Twitter corpus. Tang et al. [42] encodes sentiment information of texts (e.g., sentences and words) together with contexts of words in sentiment embeddings. They showed that sentiment embeddings consistently outperform context-based embeddings in tasks such as word-level sentiment analysis, sentence level sentiment classification and building sentiment lexicons.

Besides machine learning approaches for sentiment classification, in the literature we find lexicon-based approaches. While finding or constructing those lexicons is not always an easy task and the difficulty might vary depending on the language, these methods allow us to use a list of words (dictionary of subjective words) [22],

where each word is associated with a specific sentiment; emoticons are used the same way as well. Yadollahi et al. [48] discuss the ability to use more than one dataset to take into account multiple subjective perspectives of the word and to modify the existing dictionary in order to satisfy the topic sentiment characteristics. Asghar et al. [1] proposed to classify sentiments in reviews by combining an emoticon classifier, a modifier and negation classifier and a classifier based on the opinion lexicon SentiWordNetwork (SWNC) in a sequential, then input the text to a domain specific classifier (DSC) that takes into account the polarity of domain specific words both existing or unknown in SWNC. Such a hybrid approach consists of the combination of both machine learning and lexicon-based approaches, which can improve the results of sentiment classification. More information on related works in the area can be found in [25].

Analysis Pipeline

For our sentiment analysis task, we propose an approach that uses data mining and machine learning for the extraction of relevant tweets, then a lexicon-based sentiment classifier to calculate their polarities and classify them to negative, positive and neutral. More specifically, we provide a processing pipeline of four distinct sequential and interdependent steps. More specifically:

Collection and Preprocessing of Tweets

Tweets are collected through the Twitter API, which provides a standard way to get (a part of) the real time stream of public tweets and filter those by keywords, location, language, users, etc. We used mainly keyword-based and location-based queries. We are extracting only English language tweets and use keywords like: "Heathrow", "LHR", "noise", "annoyance", etc. to increase the chances to get relevant messages. Also, location queries were used based on Heathrow's day, evening and night level (L_{den}) noise contours in order to bound the area of interest. (This led to a bounding box of 167 km wide and 73 km long, centred around Heathrow airport to be used as location filter to Twitter API).

Those tweets go through a pre-processing phase, where we firstly remove links, numbers, emoticons and Twitter specific words, then we make all words lowercase and apply tokenisation. On those tokenised words, we correct as many errors as possible (mainly spelling errors) and then we assign part-of-speech (POS) tags and lemmatise the words in order to work with a more compact and stronger set for understanding relevance. It should be noted here that the removed parts (e.g., emoticons) are not deleted permanently but are passed to the next processing steps.

Relevance Classification

From the previous step we end up with a bag of words (including hashtags), so here we use these words to form unigrams, bigrams and hashtags as features and use tf-idf as a metric to represent tweets. Then the SVM algorithm is trained on a manually annotated sample of tweets, so as to be able to classify tweets as relevant or not. We filter the relevant tweets through a lexicon-based classifier in order to benefit from the domain knowledge (expressed through the lexicon), which assigns a relevance score to each tweet. We keep those tweets that exceed a threshold. This double classification provides better results compared with other methods, for more details please see [25].

Sentiment Analysis of Relevant Tweets

Based on the selection of relevant tweets, we proceed to classify those tweets as positive, negative or neutral. We do this by exploring various facets of the tweets and calculate different scores that represent each facet. At the end, we put together those scores in order to compute a single score per relevant tweet that would allow the system (based on this value) to classify it. In order to do this, we use three different facets: (a) emoticons (collected from tweets and labeled as positive or negative), (b) lexicon-based polarity of words (using dictionaries, where each word has been classified as positive or negative and given a score) and (c) the SentiWordNet, a dictionary where each word has been attributed at the same time a positive, a negative and a neutral score with the restriction that these scores add up to 1. The final weighted score is calculated based on the individual scores and is used for classifying the tweet.

The overall processing pipeline is depicted in Fig. 11.10 and has been published in more detail in [25].

Preliminary Results

At the time of writing this text, we already had some promising preliminary results, at least in the sense of capturing correctly the overall sentiment of the population involved, given the limitations discussed in the beginning. Although the complexity of the pipeline amplifies the errors we have in the processing, preliminary results show quite good accuracy in the classification of sentiments found in the relevant tweets. Moreover, the errors we calculate are equally distributed between the different classes, which shows that the method does not introduce any bias towards a specific class. Results can be visualised either as graphs or as localised data with the use of a map for the visual background.
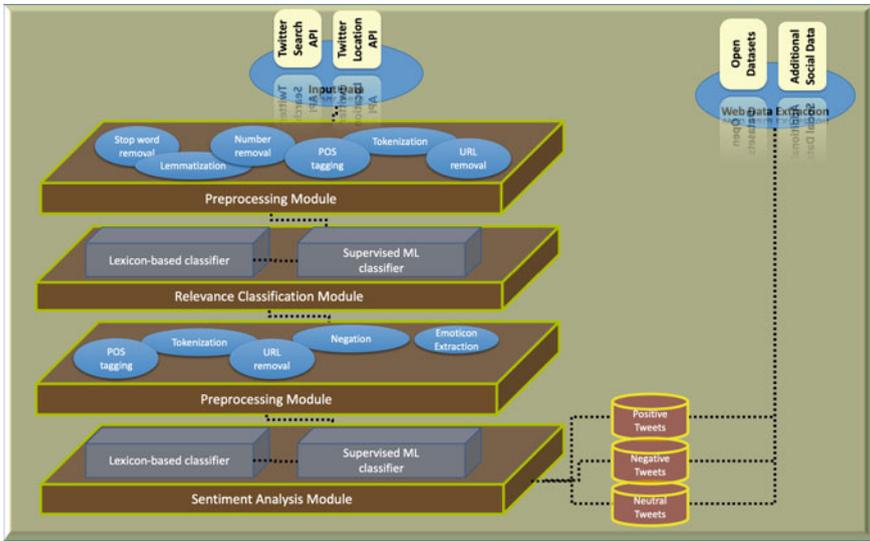
Fig. 11.10 The twitter analysis pipeline [25]

Future Work

The main effort is the large-scale application and evaluation of the proposed methodology. Given the effects of the COVID-19 pandemic on airport traffic but also on the discussions on twitter and other social media, we will rely on historical data (i.e. data recorded prior to the pandemic) to do our processing. This eases the requirements on real-time processing and allows us to apply additional methods, like embeddings, which can improve the accuracy of sentiment classification. Unlike the "bag of words" representation used by our methods so far, where the context plays a small role; these newer methods are able to detect similarities and hence classify unseen words which are similar to other words seen in the training set. Recurrent Neural Network (RNN) offers a memory factor that helps to consider the previous and the following words to better predict the sentiment of current words and hence to efficiently predict the sentiment of the whole sentence, which improves the accuracy of the classifier.

Conclusion

In this chapter, we presented how recent technological innovations could help to collect and analyse data from people who live around airports, and so improve our understanding of the adverse effect of noise on humans.

A Virtual Reality simulation made it possible to evaluate how visual settings of the aircraft (wide body, narrow body, blended wing body) and of the landscape (green park vs. urban situation) influenced sound perception. The quality of the tool has been tested, showing that 96% of the participants felt surrounded by the environment, and 78% found the virtual audio-visual environment very or extremely realistic. This application rendered highly immersive audio-visual situations, so, it has been hypothesised that it could be relevant to communicate with residents in a fair approach, showing the impact of future airport scenarios in land-use planning.

A mobile application (AnimApp) has been developed to study the impact of the audio-visual environment on sound perception and on quality of life around airports. The method of experience sampling has been chosen, because it captures subjective experiences as they are experienced in-situ in real life. Perceptual data on the sound as well as the visual environment are collected in addition to acoustic data. The final study will take place during spring/summer 2021 at two locations: one at Ljubljana Airport in Slovenia, the other at London Heathrow Airport in the UK with the aim of convincing more than 60 participants in each site. This experiment will probably suffer from the reduction of air traffic due to the COVID-19 pandemic, but this application can be also used in the future to collect valuable perceptual data synchronised with acoustic ones with more participants.

Finally, using social media as a means to survey people's opinions on various subjects, including quality of life and issues of noise around airports, seems to be promising and produces credible results. The process gives us insights based on existing online discussions and based on complex learning pipelines; it discovers, classifies and localises the opinions of the users. The complexity of the current processing architectures is significant but the results produced so far are promising for the future. Being able to combine those data with additional offline data and data from multiple sources (e.g. other opinion sites, land use, etc.) could improve the quality of the insights provided into people's responses to aircraft noise, and, thus, and allow to further refine the process of aviation noise management in airport regions.

References

1. Asghar MZ, Khan A, Ahmad S, Qasim M, Khan IA (2017) Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PLOS One 12(2):e0171649
2. Axelsson O, Nilsson M, Berglund B (2010) A principal component model of soundscape perception. J Acoust Soc Am 128(5):2836–2846
3. Bolger N, Davis A, Rafaeli E (2003) Diary methods: capturing life as it is lived. Annu Rev Psychol 54:579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030
4. Brambilla G, Maffei L (2006) Responses to noise in urban parks and in rural quiet areas. Acta Acustica united with Acustica 92(6):881–886
5. Cain R, Jennings P, Poxon J (2013) The development and application of the emotional dimensions of a soundscape. Appl Acoust 74:232–239
6. Cortes C, Vapnik V (1995) Support-vector networks. Machine Learn 20(3):273–297

7. Craig A, Moore D, Knox D (2017) Experience sampling: assessing urban soundscapes using in-situ participatory methods. Appl Acoust 117:227–235. https://doi.org/10.1016/j.apacoust.2016.05.026

8. Dedieu R, Lavandier C, Aalmoes R, Legriffon I, Boullet I (2021) Studies on virtual and auralisation tools. ANIMA Deliverable D3.10

9. Fiebig A, Sottek R (2015) Contribution of peak events to overall loudness. Acta Acust united Ac 101:1116–1129

10. Fields JM, De Jong RG, Gjestland T, Flindell IH, Job RFS, Kurra S, Lercher P, Vallet M, Yano T, Guski R, Felscher-Suhr U, Schumer R (2001) Standardised general-purpose noise response questions for community noise surveys: research and a recommendation. J Sound Vib 242(4):641–679. Last visited on the 26th of March. https://www.sciencedirect.com/science/article/pii/S0022460X00933844

11. Fujiwara D, Lawton RN, MacKerron G (2017) Experience sampling in and around airports. Momentary subjective wellbeing, airports, and aviation noise in England. Transp Res Part D: Transp Environ 56:43–54. ISSN 1361–9209. https://doi.org/10.1016/j.trd.2017.07.015. (https://www.sciencedirect.com/science/article/pii/S136192091530256X)

12. Gille L-A, Marquis-Favre C (2015) Physical and perceptual characterisation of aircraft noise to better assess noise annoyance. In: Proceedings internoise, San Francisco, USA

13. Guski R (1999) Personal and social variables as co-determinants of noise annoyance. Noise Health 3:45–56

14. Guski R (2017) The increase of aircraft noise annoyance in communities. Causes and consequences, Keynote, Proc ICBEN 2017, Zurich, Switzerland, ID 4164, (2017) Available from http://www.icben.org/2017/ICBEN%202017%20Papers/Keynote04_Guski_4164.pdf

15. Guski R, Schreckenberg D, Schuemer R (2017) WHO environmental noise guidelines for the European Region: a systematic review on environmental noise and annoyance. Int J Environ Res Public Health 14(12):1539. https://doi.org/10.3390/ijerph14121539

16. Hutto C, Gilbert E, Vader A (2014) Parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the international AAAI conference on web and social media. vol 8

17. International Wellbeing Group, Personal Wellbeing Index: 5th Edition (2013). Available from http://www.acqol.com.au/uploads/pwi-a/pwi-a-english.pdf (May 2019)

18. ISO 12913–1 (2014) Acoustics—Soundscape—Part 1: definition and conceptual framework. Geneva, Switzerland, International Organisation for Standardisation (ISO)

19. ISO 12913–2 (2018) Acoustics—Soundscape—Part 2: data collection and reporting requirements. Geneva, Switzerland, International Organisation for Standardisation (ISO)

20. Jianqiang Z, Xiaolin G, Xuejun Z (2018) Deep convolution neural networks for twitter sentiment analysis. IEEE Access 6:23253–23260. https://doi.org/10.1109/ACCESS.2017.2776930

21. Kouloumpis E, Wilson T, Moore J (2011) Twitter sentiment analysis: the good the bad and the OMG!. In: Proceedings of the international AAAI conference on web and social media. vol 5.

22. Liu B, Hu M, Cheng J (2005) Opinion observer: analysing and comparing opinions on the web. In: Proceedings of the 14th international conference on WorldWideWeb. pp 342–351

23. Loughran T, Mc DB (2011) When is a liability not a liability? textual analysis, dictionaries, and 10-ks. J Financ 66(1):35–65

24. Maisonneuve N, Matthias N (2010) Participatory noise pollution monitoring using mobile phones. Information Polity 15:51–71

25. Meddeb I, Lavandier C, Kotzinos D (2020) Using twitter streams for opinion mining: a case study on airport noise. Springer Series "Communications in Computer and Information Science". 1197:145–160

26. More SR (2011) Aircraft noise characteristics and metrics. Ph.D. Thesis, Purdue University, West Lafayette, USA

27. Ouyang X, Zhou P, Li CH, Liu L (2015) Sentiment analysis using convolutional neural network. In: IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing. Pervasive Intelligence and Computing, pp 2359–2364. https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.349

28. Page JA, Hogdon KK, Hunte RP, Davis DE, Gaugler TA, Downs R et al. (2019) Quiet supersonic flights 2018 (QSF18) test: galveston, texas risk reduction for future community testing with a low-boom flight demonstration vehicle, NASA, Hapton, VA (NASA/CR–2020–220589/Volume I), (2020). Last visited on the 26th of March 2021. https://ntrs.nasa.gov/search.jsp?R=20200003223

29. Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. LREc 10:1320–1326

30. Paté A, Lavandier C, Minard A, Le Griffon I (2017) Perceived unpleasantness of aircraft flyover noise: influence of temporal parameters. Acta Acust Acust 103:34–47. https://doi.org/10.3813/AAA.919031

31. Pheasant R, Horoshenkov K, Watts G, Barrett B (2008) The acoustic and visual factors influencing the construction of tranquil space in urban and rural environments tranquil spaces-quiet places? J Acoust Soc Am 123(3):1446–1457

32. Preis A, Hafke-Dys H, Szychowska M, Kociński J, Felcyn J (2016) Audiovisual interaction of environmental noise. Noise Cont Eng J 64:36–45

33. Psychiatric Research Unit, WHO Collaborating Centre in Mental Health. WHO (Five) Well-Being Index, (1998) Available from https://www.psykiatri-regionh.dk/who-5/Documents/WHO-5%20questionaire%20-%20English.pdf Last visited on the 26th of March 2021

34. Radicchi A, Henckel D, Memmel M (2018) Citizens as smart, active sensors for a quiet and just city. The case of the 'open source soundscapes' approach to identify, assess and plan 'everyday quiet areas' in cities. Noise Mapping J 5:1–20

35. Read J (2005) Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In: Proceedings of the ACL student research workshop, pp 43–48

36. Ren Y, Zhang Y, Zhang M, Ji D (2016) Context-sensitive twitter sentiment classification using neural network. In: Proceedings of the AAAI conference on artificial intelligence. vol 30

37. Ricciardi P, Delaitre P, Lavandier C, Torchia F, Aumond P (2015) Sound quality indicators for urban places in Paris cross-validated by Milan data. J Acoust Soc Am 138(4):2337–2348

38. Rizzi SA, Sullivan BM, Aumann AR (2008) Recent developments in aircraft flyover noise simulation at NASA langley research center. NATO RTO Specialist Meeting AVT-158 on Noise Issues Associated with Gas Turbine Powered Military Vehicles, 13–17 Oct, Montreal, Canada

39. Stat Counter Global Stats, Mobile Operating System Market Share in Europe (2019) Online Available from: http://gs.statcounter.com/os-market-share/mobile/europe (March 2021)

40. Steffens J, Guastavino C (2015) Trend effects in momentary and retrospective soundscape judgments. Acta Acust United Acust 101:713–722

41. Susini P, McAdams S, Smith BK (2002) Global and continuous loudness estimation of time-varying levels. Acta Acust united Ac 88:536–548

42. Tang D, Wei F, Qin B, Yang N, Liu T, Zhou M (2015) Sentiment embeddings with applications to sentiment analysis. IEEE Trans Knowl Data Eng 28(2):496–509

43. Van Renterghem T, Botteldooren D (2016) View on outdoor vegetation reduces noise annoyance for dwellers near busy roads. Landsc Urb Plan 148:203–215

44. Västfjäll D (2004) The 'end effect' in retrospective sound quality evaluation. Acoust Sci Technol 25:170–172

45. Viollon S, Lavandier C, Drake C (2002) Influence of visual setting on sound ratings in an urban environment. Appl Acoust 63(5):493–511

46. WHO Environmental Noise Guidelines for the European Region (2018) ISBN 978 92 890 5356 3. Last visited on the 26th of March 2021. https://www.euro.who.int/en/health-topics/environment-and-health/noise/publications/2018/environmental-noise-guidelines-for-the-european-region-2018

47. Wilson T, Wiebe J, Hoffmann P (2005) Recognising contextual polarity in phrase-level senti-
    ment analysis. In: Proceedings of human language technology conference and conference on
    empirical methods in natural language processing, pp 347–354
48. Yadollahi A, Shahraki AG, Zaiane OR (2017) Current state of text sentiment analysis from
    opinion to emotion mining. ACM Comput Surv (CSUR) 50(2):1–33