PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Auralization of aircraft noise in an urban environment

Roalt AALMOES[1]; Maurits VAN DER VEEN[2]; Henk LANIA;[3]

[1, 3] Royal Netherlands Aerospace Centre, The Netherlands

[2] Utrecht University, The Netherlands

ABSTRACT

Realistic simulation of aircraft fly-over using auralization and visualisation techniques help in the research towards more silent aircraft types and/or operations. Up until recently most auralizations of aircraft were created in a virtual countryside, only taking the direct path and ground reflection into account. In an urban environment, reflections and absorption of buildings must be considered as well as different ambient sound levels. This research effort aims at creating an auralization of a VTOL aircraft flying through an urban environment. A dynamic 3D model of the urban environment is constructed using publicly available city data. This data is combined with recorded video data to match visual and audio occlusion of the aircraft by buildings. Reflection surfaces are used to determine mirrored noise sources, which are spatially added to the 3D buildings. At the observer the sound from the different paths is received and presented as binaural (directional) sound together with a realistic visualisation in Virtual Reality using the Oculus Rift.

Keywords: Aircraft noise, Virtual Reality, Urban environment.

1. INTRODUCTION

Noise annoyance from aircraft is one of the most adverse effects on communities living around airports (1). Scientific research on noise annoyance helps in finding and addressing the main causes of annoyance, either in sound level, sound characteristics, or non-acoustical factors. Although this research can take place around airports with affected communities, a controlled environment is preferable for receiving measurable data. For this purpose, the Royal Netherlands Aerospace Centre (NLR) developed, based on prior research by NASA Langley (2), the Virtual Community Noise Simulator (VCNS), to conduct research and inform communities on the effects of (measures on) aircraft noise for their environment (3, 4). But until now, this research does not take new modes of air transport into account, such as drones and Vertical Take-Off and Landing (VTOL) aircraft.

Figure 1 – Camera rig for panoramic recording

[1] Roalt.Aalmoes@nlr.nl
[2] mauritsvanderveen@mail.com
[3] Henk.Lania@nlr.nl

Urban Air Mobility (5, 6, 7) (UAM) is an upcoming air transport concept and expected to have a significant influence on people's life, for the better or the worse: a new fleet of electrical powered aircraft including package-delivery drones and VTOL aircraft may reduce road congestion, improve travel or transport times for people, but may also have an adverse effect with respect to noise or visual impact to people living in the city. Noise calculations to evaluate noise impact of UAM concepts have been done before by Arntzen et al (8), but cannot be used to perform annoyance research. To do such research using a simulated environment, the variables have changed: Up until recently, most auralizations of aircraft were created in a virtual countryside, only taking the direct path and ground reflection into account.

In an urban environment, the following factors should be addressed as well: First, sound reflections and absorption of buildings must be considered as well as different ambient sound levels. Second, a visualisation within a city requires a sophisticated model where buildings or other objects may at one moment occlude an aircraft, but can at another moment be in the background of a moving aircraft in front. Third, as aircraft maybe closer by, shadow casting effects may have a visual impact on communities and may contribute to or even amplify the existing (negative) attitude towards (nearby) aircraft. Concluding, significant changes are needed to simulate urban air mobility concepts compared to traditional aircraft flyover cases.

2. PROPOSAL FOR AN URBAN ENVIRONMENTAL SIMULATOR

To understand the requirements for an urban environmental simulator, first the existing VCNS simulation setup must be described. In a traditional aircraft fly-over scenario, a location is filmed using a 360-degree camera. This can be done using a combination of cameras in a specific 360 degrees camera rig, such as done at NLR (See Figure 1) or it can be done with a dedicated panoramic camera. For the first approach, dedicated software is required to stitch the video together to generate a panoramic 360-degree video. For the second approach, the stitching is done automatically by either the hardware or the associated software. For also recording the ambient sounds, a first order microphone is used to record directional sounds. This is sufficient for users of the simulation to get an impression of the direction for certain ambient sounds. An audio cue is used at the beginning of the recording to synchronize audio and video (4) later on in the process. The aircraft fly-over is later added to the recording by either creating a (fully-controlled) auralization, or by making use of a recording of an aircraft and adjusting sound level and atmospheric absorption based on the difference in distance between recorded and play-back location. In these simulations, there is always a clear line-of-sight and sound and visuals are only obscured by objects near the horizon, such as houses and trees, but they do not significantly influence the audio path between source and observer. So for the audio, no measures need to be taken, and for the video, only a simple masking-filter is used to indicate whether the aircraft should be displayed before (white) or after (black) the video background recording, see Figure 2. The VCNS software can superimpose the mask on an artificially added 3D aircraft object so only (part of) the aircraft is shown if it is above the horizon. Finally, everything is combined and by making use of an Oculus Rift Virtual Reality headset and binaural audio reproduction, a simulation can be played in an immersive environment.

For simulation of an urban city centre, the buildings in the environment play a more dominant role in the simulation, with characteristics that have influence on the perception by the observer. Except for the usual cases where aircraft are either both visible and audible or not, aircraft may be audible but not visible, or only audible by one or more reflections. Another difference is that some of the proposed vehicles in the UAM concept are electric multi-copter aircraft with different noise characteristics than

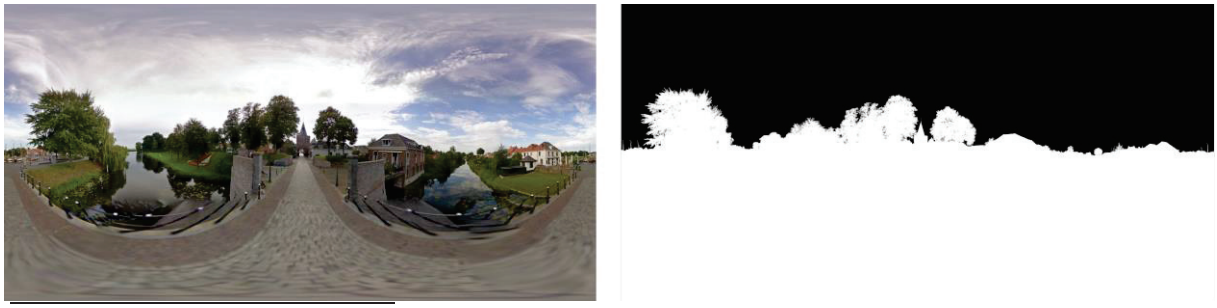

Figure 2 – Left, a screenshot of a panoramic video, and right, a masking image to filter foreground (white) from background (black)

traditional fixed-wing or helicopter aircraft. Finally, ambient sound levels are different and generally louder, which masks aircraft sounds, but also causes a different attitude by the observer for expected noise annoyance at that location.

2.1 Visualisation

For the visualisation, a more sophisticated 3D model has been developed and be used as overlay for the recorded panoramic video. In this paper, a summary is given of this work by van der Veen (9). Creating such a model by hand is possible but a cumbersome task. The following tasks are needed to generate such a model automatically:

1. **Initial model generation**: Publicly available height data is used to reconstruct an initial 3D city model of the recorded location. This research will focus on the Netherlands, for which high resolution aerial LiDAR data is publicly available (10). A virtual camera is placed at the correct location in the 3D city model by using manually estimated GPS coordinates.
2. **Camera pose optimization**: To use the 3D model for augmenting a panoramic video, the estimated global position and rotation of the camera need to be optimized. This should be done automatically.
3. **Model refinement**: Once the 3D city model is aligned with the panoramic video, the edges of the model will be adjusted so that the visible edges in the video closely match the edges of the 3D model. This is necessary, because large differences between the video and the 3D model will be clearly noticeable, especially when rendering occluded virtual objects.
4. **Depth video generation**: The initial 3D city model will not contain all objects that appear in the video, because this requires a perfectly up-to-date, high resolution data source and perfect reconstruction from the data. Objects like trees, street lights, cars, and pedestrians will typically be missing. In this thesis, those objects will be referred to as urban objects. Our system should automatically detect these objects and estimate their depth in each video frame. This will be used to create a depth video.

The whole process of generating and matching a 3D city model and depth video should be automated as much as possible. Some manual modelling is acceptable for locations where the height data is not up-to-date. The final 3D city model with optimized camera pose and the depth video will be used to augment the panoramic video by adding virtual objects. In this case, the virtual objects will be drones. We assume that the video is recorded in an outdoor environment with sunny weather. The virtual objects will be rendered with correct occlusion and they will cast and receive shadows. The system that will be developed is called an Augmented Virtual Environment (AVE). An AVE is defined as a fusion of dynamic imagery with 3D models (11). In our AVE, the camera will be stationary.

2.1.1 Proposed model

Our method automatically generates a 3D city model and depth video, which are used to create an

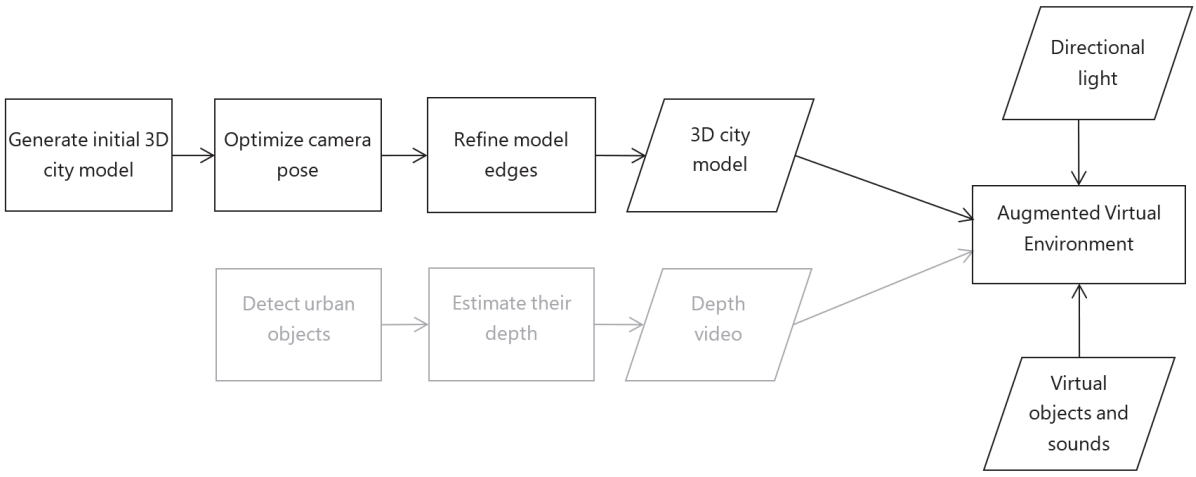

Figure 3 – Proposed method for Augmented Virtual Environment

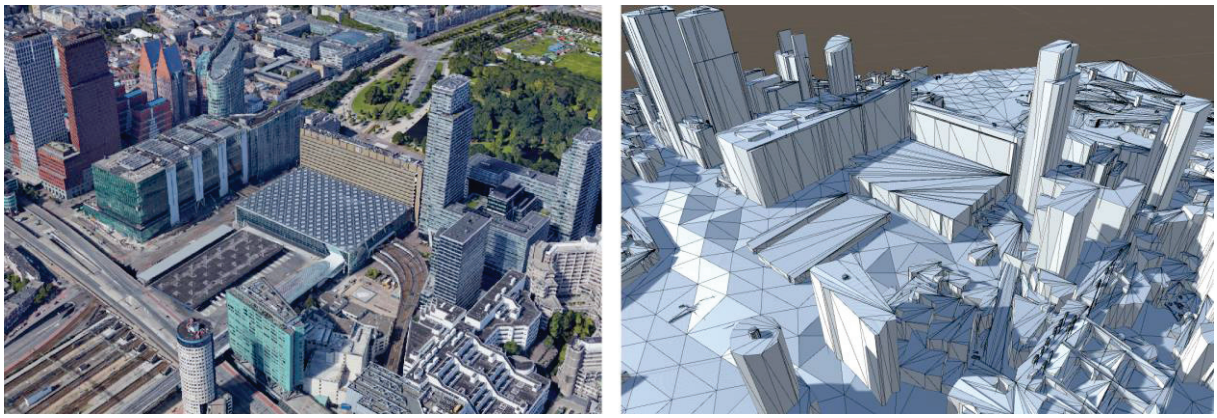

Figure 4 – On the left an existing Google Earth 3D model of the city centre of The Hague, and on the right the same location but recreated using our 3D city model.

AVE, given a 360-degree video, estimated GPS coordinates, estimated camera height, aerial LiDAR data of the recorded location and the altitude and azimuth of the sun, see Figure 3. It is implemented in the Unity gaming engine[5]. The automatic generation of a 3D city model and depth video has four components. The first component reconstructs a 3D city model from publicly available high resolution aerial LiDAR data. Using high resolution height data makes it possible to reconstruct detailed 3D structures. Having detail in these initial 3D structures makes the next steps easier, because removing detail is easier than adding detail. The data is first segmented by using a region growing method, as proposed by Vosselman (12). This method is used because it is the most versatile segmentation method, it works for any roof shape. It is also easy to fine tune the segmentation process, by changing the region growing criteria. The boundaries of the segments are traced similarly to how it was done by Sampath and Shan (13). Then, the boundaries are simplified using a line fitting method, which was also done by Poullis and You (14). This method works for any boundary shape and it is easy to change how much detail should be preserved. Finally, the simplified boundaries are used to create a 3D mesh from each segment, see Figure 4. In some cases, buildings will have to be manually added or removed from the 3D city model or the height data, because the height data is not always up-to-date. A model to take moving urban objects, such as cars and pedestrians, into account was developed as well but considered outside the scope for this article.

The camera pose optimization is based on the method by Hofmann et al. (15). This naive method is sufficient, because in our case the initial camera pose will be relatively accurate. Furthermore, sunny weather is assumed, which makes it easy to extract the sky from a video frame. The camera pose is optimized by maximizing the overlap of the sky in the video and the sky of the 3D city model. This is done with a local search on the camera parameters, using a best improvement heuristic. It should find the correct camera pose, assuming that the initial estimate camera pose is sufficiently accurate.

The next component refines the edges of the 3D model, given the initial 3D city model, improved camera pose and panoramic video. For this, a method is used which is similar to how Pu and Vosselman (16) match model edges with images, see Figure 5.

Finally, virtual objects are added to the video by using the generated and matched 3D city model and the depth video. The city model handles occlusions of buildings, receives shadows from virtual objects and casts shadows on virtual objects, while keeping the city model invisible. The virtual objects are lit by a single directional light source, which simulates the sun.

2.2 Auralization

The generated 3D model for the visualisation can be reused for the auralization as well, and therefore, serves a second purpose: it can be used as a model to calculate sound reflections and absorptions. In this research, we focus on absorption and (combinations of) single reflections only: these effects have the most impact for the perception of aircraft noise in urban environments, see Figure 6. But with this 3D model, this can later be expanded by also considering sound diffraction from edges. Surface materials of this model are not known, so an estimate must be made here with respect to

[5] https://unity.com/

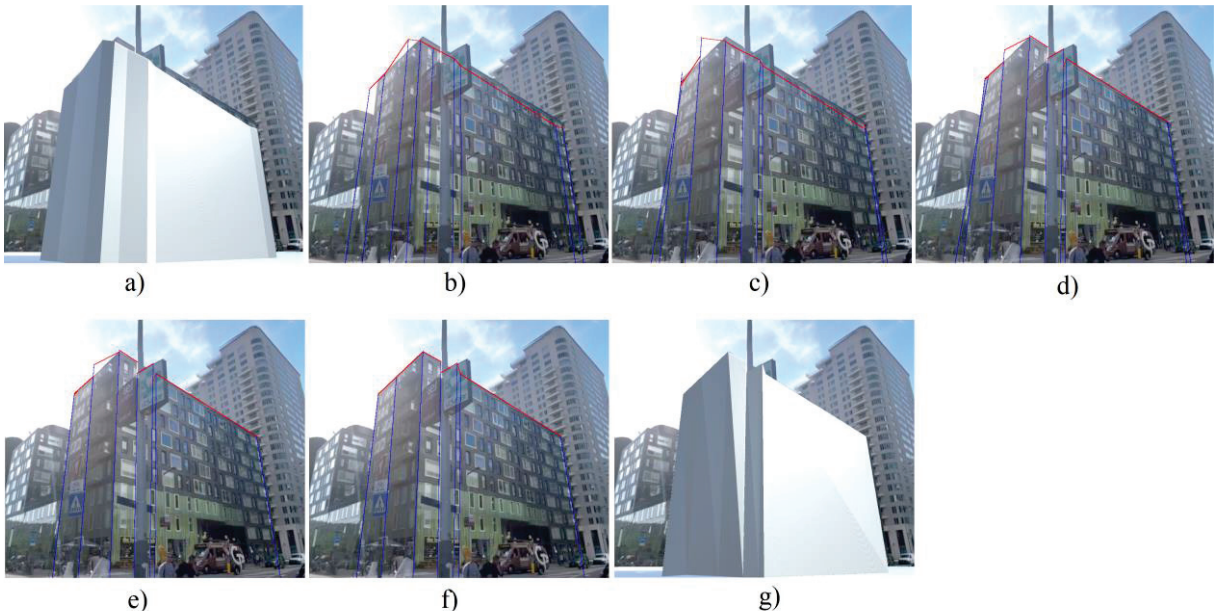

Figure 5 – Edge refinement process of a building (blue lines are the wall edges and red lines the roof edges), with: a) The generated 3D mesh with optimized camera pose b) Projection of mesh edges to 2D c) optimization of wall edges d) Optimization of roof edges e) Removal of intersecting wall edges f) Smoothing of roof edges g) The refined 3D mesh

reflection and scattering of sound. This set up will serve as a proof-of-concept for the approach described and allows real-time audio reproduction for the observer. For an unobstructed sound reproduction between source and observer, an approach can be followed as earlier demonstrated in Aalmoes (8): Aircraft sound is generated at the source, based upon earlier recordings. In this case, a VTOL recording found on YouTube was used as source and is used, until better sound sources are available. Optionally, atmospheric absorption is applied based upon the distance between source and observer and the atmospheric conditions at the given location. Additionally, direct path and ground reflection are calculated and combined. For the ground reflection, the type of ground (e.g. concrete or grass) can be taken into account.

For obstructed paths from source to observer, the facades of the 3D city models are evaluated to calculate if their reflected path, using the facade as mirror, can reach the observer. If so, virtual sound sources are added in the 3D environment on the location of the reflection. A frequency-specific filter will be added to this added sound source to correct for absorption of the building. At this moment, the material of the building is not further determined, but considered to be concrete, with limited absorption capabilities. Improved 3D models can help in the future to provide correct reflection characteristics. Finally, binaural reproduction in combination with motion sensors from the virtual reality headset is used to provide the observer with a directional sound field.

For adding ambient sounds, first order 3D audio recordings, using a Zoom H2n or a Core Sound TetraMic microphone, are made at the same time and location while recording the video. These audio recordings are synchronized with the video and added using a spatial replay module to give realistic ambient sound atmosphere. The combination of directional sound of the virtual aircraft, the ambient sound from the environment, the virtual model of the VTOL aircraft, and the depth-enhanced panoramic video receiving cast shadows, results in an environmental simulation for UAM concepts in Virtual Reality.

3. RESULTS

For the visualisation, both an objective study and a subjective study were conducted. Auralization was not yet available and therefore not included in the study. The goal of the objective study was to evaluate the effectiveness of the automated generation of a 3D city model. Results from the user study are not discussed here, but can be found in (9).

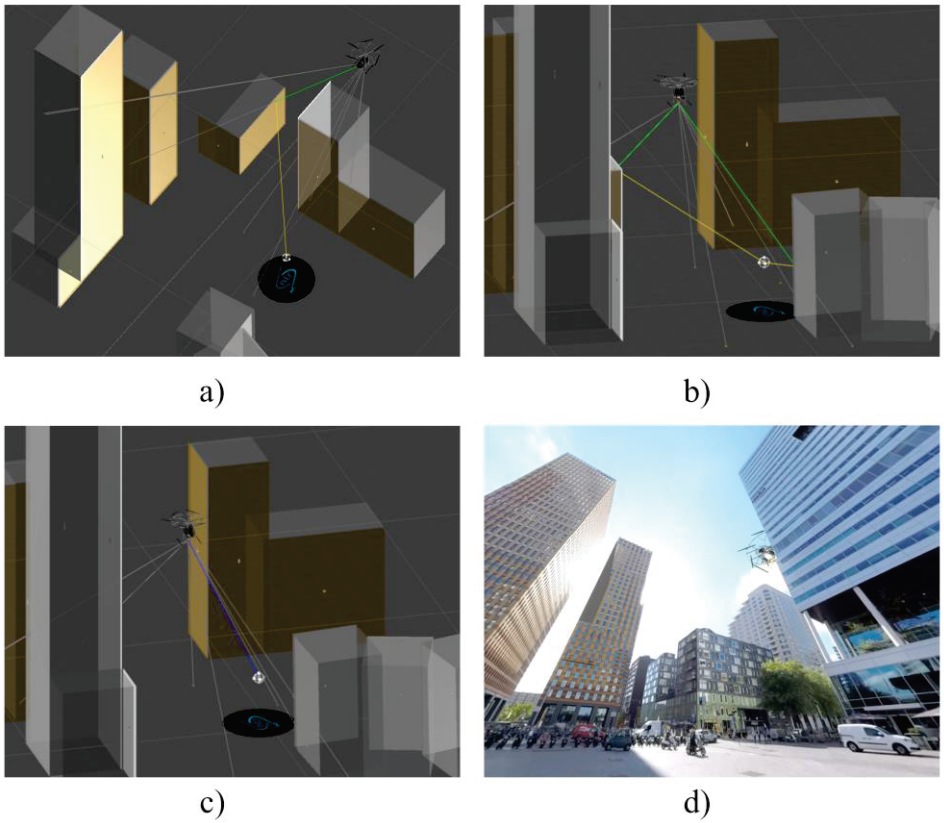

Figure 6 – Phases of reflection of an approaching VTOL aircraft towards the observer (white object) with a) a single reflected sound path where green is the path before the reflection, and yellow the path after the reflection, b) A double reflection path, c) a direct sound path from source to observer (in blue), and d) the view by the observer in the virtual reality headset when there is a direct path.

3.1 Camera pose optimization

The camera pose optimization was evaluated by running it multiple times on each test video. The initial generated 3D city models were used, so without refined edges, because the edge refinement is done in a later stage. The starting position and rotation of the camera was varied for each run. This was done to test whether our camera pose estimation is able to find the correct camera pose, given a good, bad, decent or terrible estimation. The starting positions and rotations were obtained by placing the camera at different distances and angles away from the ground truth. A distance of 5 meters and an angle offset of 3 degrees were used as a good estimate. This was 10 meters and 5 degrees for a decent estimate. A bad estimate was 20 meters and 8 degrees away from the ground truth. Finally, a terrible estimation was simulated by a distance of 25 meters and an angle offset of 10 degrees. Each category was evaluated by using three positions on a circle around the ground truth. Since there were four test locations, each estimate pose category was evaluated using 3x2x4 = 24 different input poses. After each run, the difference between the found camera pose and ground truth camera pose was computed.

A found camera pose was considered correct, if the difference between the ground truth was less than the used step size. A step size of 0.5 meter for the position and 0.25 degree for the rotation was used. In some cases, the found camera poses was close to the ground truth (less than 2 meters away) and had a slightly higher overlap score than the ground truth. This is either due to imperfections in the sky segmentation or because the ground truth that was used is not the actual ground truth. These found camera poses were considered semi-correct. A semi-correct camera pose means that it is only slightly off from the ground truth. It is likely that these poses will not cause problems when refining the model edges in the next processing step, so they will yield the same final result as a correct camera pose.

Looking at all test locations together, the ground truth pose was found for only half of the locations, given a good estimate (Figure 7). The found poses that were more than two meters away from the ground truth position, were always a local optimum. The overlap scores of those poses were significantly lower than the score of the ground truth poses. A semi-correct camera pose was found for

three of the four locations, given a good estimate. A decent estimate was sufficient for two of the four locations. For one of the four videos, the optimized coordinates were more than 20 meters away from the estimated coordinates. This was due to an error in the sky segmentation. A large white building was lit by the sun, which caused it to be segmented as part of the sky.

3.2 Model refinement

The model refinement was run on each generated 3D model, given the panoramic videos and optimized camera poses. The camera poses were optimized from a good estimation of the GPS coordinates. After running the model refinement, the error of each refined model was computed. This was based on ground truth segmentations of a single frame of each video. Since we are evaluating the model refinement, which only models buildings and large structures, other objects should be ignored. Pixels in the bottom half of the images were also ignored, because those are typically terrain pixels and our method does not refine the edges of the terrain mesh. Each pixel in the image was compared to the same pixel of its ground truth segmentation image. If a pixel in the model view was a sky pixel, but the same pixel was a building pixel in the segmentation image, or vice versa, then it was counted as an error. The percentage of wrong pixels was then computed by dividing the number of error pixels by the total number of evaluated pixels. This percentage is a measure of the error of the refined 3D model. Around 3.5 million pixels were evaluated for each location, which is slightly less than half the number of total pixels in a video frame.

The four evaluated locations had error rates of 1.4%, 1.8%, 4.9%, and 9.8%. For the latter two locations, sun light and mismatches in the horizon (possibly caused by stitching errors for the generation of the panoramic video) resulted in errors, and for one building the contrasting windows were mistakenly identified as the building edge.

4. CONCLUSION

This research presents an approach to make a demonstrator for an urban environment with simulated aircraft sound, in particular low-flying drones and VTOL aircraft. Such a demonstrator can help with further research on UAM concepts and their impact on communities, and define mitigation measures if needed. It can also define research to build cities to reduce the impact of this new concept, as done earlier for impact with aircraft noise (17).

A 3D city model is necessary to do the auralization, when taking reflections and absorption into account. It also helps in combination with 360-degree video to create an immersive environment, including added (artificial) aircraft and their cast shadow effects. With the proper aerial LiDAR data, this model can be created automatically, and the camera pose position of the recorded video can also be determined automatically. After edge refinement is applied, a depth model is available and allows aircraft to be visualised before, between, or behind the existing buildings. This model can also be used to take absorption and reflection for sound sources into account, as demonstrated here with a simple sound reflection model. An objective study on the quality of the model indicates that automatic generation is possible, if quality of the panoramic stitching and influence of sunlight is taken into consideration.

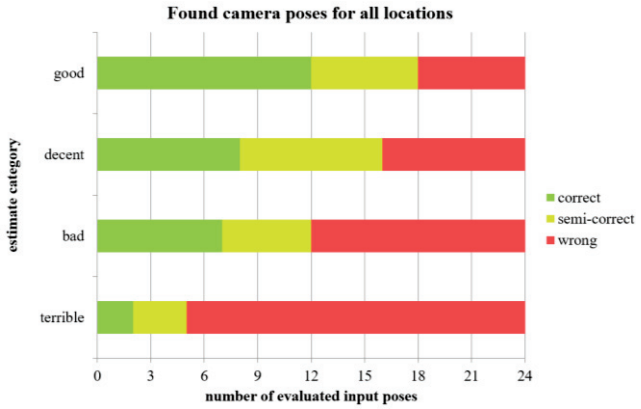

Figure 7 – Number of correct, semi-correct and wrong found camera poses for all locations together, when running

Improvements for the visualisation can be made by taking object recognition into account by classifying pedestrian, cars, but also trees, traffic signs and poles and determining whether they are in front or behind the virtual objects. Also, future georeferenced data will become available that can be used to determine and classify the environment. This can also be used to improve the auralization by making use of (reflection) characteristics of materials from georeferenced data. Finally, adjustments of parameters and selecting the right locations, time of day, and weather conditions for 360 recorded areas can determine the success of this automated method.

For the auralization, a refined sound source model for aircraft in the UAM concept is needed, where operational procedures and directivity are also considered. Computational sound propagation models, as described by Hornikx (18), would help with a more advanced simulation in an urban city but he also describes the challenges still to overcome with these models.

REFERENCES

1. ICAO Website, retrieved at 9 May 2018 from: https://www.icao.int/environmental-protection/Pages/noise.aspx
2. Stephen A. Rizzi, Brenda M. Sullivan, Aric R. Aumann, 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. 2008, Montreal, Canada.
3. Michael Arntzen, *Aircraft noise calculation and synthesis in a non-standard atmosphere*, Doctoral Dissertation, ISBN-978-94-6259-464-7, TU Delft, 2014.
4. Roalt Aalmoes, Merlijn den Boer, Henk Veerbeek, Virtual Reality Aircraft Noise Simulation for Community Engagement, *INTERNOISE 2018*, Chicago, IL, USA.
5. Porsche Consulting. (2018). *The Future of Vertical Mobility*. Stuttgart, Germany: Porsche Consulting.
6. Air taxis: we have lift-off… (2019, May 17), *The Guardian*, Retrieved from https://www.theguardian.com/technology/2018/mar/04/air-taxis-we-have-lift-off-airbus-vahana-e hang-volocopter-uber-elevate-lilium
7. NASA. (2017, November 8). *NASA embraces Urban Air Mobility, calls for market study*. Retrieved from NASA website: https://www.nasa.gov/aero/nasa-embraces-urban-air-mobility
8. M. Arntzen, R. Aalmoes, F.J.L. Bussink, E. Sunil, J. Hoekstra, Noise computation for future urban air traffic systems, *INTERNOISE2015*, San Francisco, CA, USA.
9. Maurits R. van der Veen, *Automatic Generation of 3D City Models and Depth Videos for Augmented Virtual Environments*, Master Thesis, University Utrecht, May 6, 2019.
10. PDOK - AHN3 downloads. (2018, October 2). Retrieved from https://downloads.pdok.nl/ahn3-downloadpage/
11. Jian, H., Liao, J., Fan, X., & Xue, Z. (2017). Augmented virtual environment: fusion of realtime video and 3D models in the digital earth system. *International Journal of Digital Earth*, 10(12), 1177-1196.
12. Vosselman, G. (1999). Building reconstruction using planar faces in very high density height data. *International Archives of Photogrammetry and Remote Sensing*, 32(3), 87-94.
13. Sampath, A., & Shan, J. (2007). Building boundary tracing and regularization from airborne LiDAR point clouds. *Photogrammetric Engineering & Remote Sensing*, 73(7), 805-812.
14. Poullis, C., & You, S. (2009). Automatic reconstruction of cities from remote sensor data. *Computer Vision and Pattern Recognition*, 2775-2782.
15. Hofmann, S., Eggert, D., & Brenner, C. (2014). Skyline matching based camera orientation from images and mobile mapping point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(5), 181.
16. Pu, S., & Vosselman, G. (2009). Building facade reconstruction by fusing terrestrial laser points and images. *Sensors*, 9(6), 4525-4542.
17. M.C. Lugten, *Tranquillity by design*, Doctoral Dissertation, Clare College, Department of Architecture and History of Art, Martin Centre for Architectural and Urban Studies, 12 December 2018.
18. M. Hornikx, Ten questions concerning computational urban acoustics, *Building and Environment*, Elsevier, Volume 106, September 2016, Pages 409-421.