Ljudsyntes av trafikbuller

Undersökning av möjligheter att automatisera ljudsyntes av trafikerade stadsmiljöer

Kandidatarbete vid institutionen för Arkitektur och Samhällsbyggnadsteknik

Ebbe Ledin
Max Persson
Andreas Törnkvist

ACEX11-VT26-61A

INSTITUTIONEN FÖR ARKITEKTUR OCH SAMHÄLLSBYGGNADSTEKNIK
Avdelning för Teknisk Akustik

CHALMERS TEKNISKA HÖGSKOLA
Göteborg, Sverige 2026
www.chalmers.se

Sammandrag

Detta kandidatarbete syftade till att undersöka olika möjligheter för att ljudsyntetisera trafikbuller från olika trafiksituationer i stadsmiljö. Under projektets gång undersöktes skillnader mellan syntetiskt skapade ljud, genom auralisering samt generativ AI, med verkliga inspelningar av stadstrafik i Göteborg. Omgivningsbuller från trafik i stadsmiljöer är idag ett växande problem och kan kopplas till flertalet hälsorisker. Behovet av ett verktyg för att kunna förutsäga den resulterande ljudmiljön i framtida stadsplanering är därav ett stort och kvaliteten av möjliga metoder för att ta sig dit av stor vikt.

I detta projekt undersöktes främst auralisering av trafikbuller från olika trafiksituationer som en potentiell metod att använda för framtida modeller, men även offentligt tillgängliga AI-modeller och deras nuvarande kapacitet att utföra samma arbete. Arbetet utfördes genom inspelningar vid diverse trafikerade områden i Göteborgs innerstad, auraliseringar av dessa inspelningar samt AI-genererade versioner och slutligen ett lyssningstest med frivilliga deltagare för att perceptuellt utvärdera slutprodukterna.

I de lyssningsförsök som utfördes upplevdes de auraliserade samt AI-genererade ljuden inte realistiska i jämförelse med de verkliga ljudmiljöerna. Däremot är bedömningen att det finns stor potential i auraliseringsmetoden och ett antal felkällor och potentiella förbättringar diskuterades.

Abstract

This project aimed to assess different possibilites to synthesize the sound environments of traffic noise for different traffic situations in an urban environment. During the process of the project, different synthesized sounds were made through auralization and with generative AI, and compared to recordings of traffic noise in Gothenburg. Noise from traffic in urban environments is an increasing issue and can be contributed to a range of health problems amongst urban populations. The need for a tool that can predict the resulting sound enivronment of a planned construction project is therefore of necessity for improving future city planning.

In this project mainly auralization of traffic noise in different urban environments was examined as a potential method for future models, also evaluations of publicly available AI-models were made in order to determine their current capacity to perform similar synthesis. The work was made through recordings in a selected diverse range of traffic environments in Gothenburg's inner city, auralizations of the recordings and AI-generated versions were then evaluated through listening experiments with volunteers to perceptually evaluate the final products.

In the listening experiments that were made the auralized and AI-generated audio signals were percieved as not particularly realistic in comparison to the actual recorded sound environments. However, the final verdict determined a vast potential in these auralization methods and some sources of error and areas of improvements were recognized and discussed.

Förord

Detta kandidatarbete genomfördes vid avdelningen Teknisk Akustik genom institutionen för Arkitektur och Samhällsbyggnadsteknik på Chalmers tekniska högskola. Projektet utfördes mellan vecka 5-20 under vårterminen 2026 med en grupp bestående av tre gruppmedlemmar, varav två från civilingenjörsprogrammet Teknisk Fysik och en från civilingenjörsprogrammet Samhällsbyggnadsteknik.

Gruppen riktar ett stort tack till både vår handledare Krister Larsson och examinator Jens Forssén för sitt stora engagemang och omfattande stöd längs hela arbetet. Vi vill även tacka Sindija Franzetti från Fackspråk och kommunikation för att ha delat med sig av sina mycket givande tankar och råd gällande skrivprocessen för den slutgiltiga rapporten. Slutligen vill vi även uttrycka vår stora uppskattning till alla personer som deltog och utförde det lyssningstest som utfördes.

Ebbe Ledin
Max Persson
Andreas Törnkvist

Göteborg, maj 2026

Innehåll

1 Inledning **1**

2 Teori **2**

2.1 Auralisering 2

2.1.1 Dämpning från propagering 2

2.1.2 Markeffekt 3

2.1.3 Dopplereffekt 3

2.2 Analys av ljud-data 4

2.3 Statistisk analys 4

2.3.1 Icke-parametriska rangtester 5

2.3.2 Konfidensintervall för binomial data 6

3 Metod **6**

3.1 Ljudinspelningar 7

3.1.1 Inspelningstillfällen 9

3.2 Auralisering 11

3.2.1 Käll-ljud 11

3.2.2 Enskilda passager 12

3.2.3 Mer komplexa trafiksituationer 14

3.3 Evaluering av tillgängliga AI-modeller 14

3.4 Utvärderingstester 16

3.4.1 Lyssningstest med deltagare 16

3.4.2 Analysmetod av lyssningstest 17

4 Resultat **17**

4.1 Prestanda 17

4.2 Visualisering av ljudfiler . . . 19
4.3 Analytiskt resultat av lyssningstest . . . 21
4.3.1 Lyssningstest - Del ett . . . 21
4.3.2 Lyssningstest - Del två . . . 22
4.3.3 Lyssningstest - Del tre . . . 22

5 Diskussion **24**

5.1 Prestanda . . . 24
5.1.1 Renderingstid av auraliseringar . . . 24
5.2 Visualiserade ljudfiler . . . 25
5.3 Lyssningstest . . . 25
5.3.1 Lyssningstest - Del ett . . . 25
5.3.2 Lyssningstest - Del två . . . 26
5.3.3 Lyssningstest - Del tre . . . 27
5.3.4 Lyssningstester - Kommentarer från deltagare . . . 27
5.4 Förbättringsområden . . . 28
5.5 Slutsats . . . 29

A Data från lyssningstester **32**

B Auralisering, Implementering i Python **35**

C Metadata för använda käll-ljud **36**

1 Inledning

Skadliga nivåer av omgivningsbuller i stadsmiljöer är ett växande problem i dagens samhälle. Den dominerande källan för omgivningsbuller orsakas främst av trafik och har funnits kopplat till flera hälsorisker som bland annat försämrad livskvalitet, hjärt-kärlsjukdomar och psykisk ohälsa [1]. Problem kopplade till bullerstörningar har utöver dess påverkan på mänsklig hälsa även ekonomiska konsekvenser och kostnaderna uppgår enligt European Environmental Agency till minst 95,6 miljarder Euro årligen [2].

Sverige har lagstiftning (förordning 2004:675) som säger att kommuner med över 100 000 invånare ska, minst var femte år, kartlägga buller i form av dygnsekvivalenta mått av ljudnivå inom kommunen [3, 4]. Att mäta och kartlägga ljudnivå är alltså väl etablerat men att enbart utgå från ljudnivå visar inte nödvändigtvis hela bilden av hur människor störs av buller [5]. Att auralisera ljud är att skapa ett hörbart ljud ifrån en eller flera källor med data som kan vara simulerad, uppmätt eller syntetiskt framtagen [6]. I kontext av stadsplanering och trafik innebär det att man simulerar hela det hörbara frekvensspektrat och på så sätt skapar en bild av det ljud som man faktiskt skulle höra om man befann sig på platsen. Att implementera auralisering som ett verktyg hade skapat unika möjligheter för en förbättrad stadsplanering för att mildra bullernivåer. I dagsläget finns auraliseringsverktyg för att simulera inomhusmiljöer men liknande verktyg för att evaluera utomhusmiljöer har inte utvecklats i samma utsträckning [7]. Att utveckla verktyg för att skapa auraliseringar i utomhusmiljöer och trafik är ett område med stor potential som kan ge en bättre bild av hur buller faktiskt upplevs. Detta skulle i sin tur driva möjligheter att positivt påverka människors hälsa och den allmänna stadsmiljön.

Detta projekt har grundat sig i de principer för auralisering som redogörs för i projektet *LISTEN Auralization of Urban Soundscapes* [7] från 2011. De metoder som användes för att skapa auraliseringar i *LISTEN* inkluderar dämpning i propageringsmedium, markeffekt, dopplereffekt m.m. Dessa metoder beskrivs vidare i korthet i teori-delen (kap. 2.1) av denna rapport. Metoden togs fram för att visa potentialen av auralisering i framtida stadsplanering med hänsyn till hur den akustiska miljön i städer blivit ett allt växande problem och intresseområde. *LISTEN* var framgångsrika i att ta fram och implementera auraliseringar av majoriteten av de studerade miljöfaktorerna. Dock saknades det vid tillfället för projektet beräkningskapacitet att genomföra mer komplexa situationer vilket är området detta projekt ska undersöka.

Detta projekt syftade till att bygga vidare på principerna i *LISTEN* för att på ett mer automatiserat vis kunna modellera en representativ uppskattning av en upplevd ljudmiljö och dessutom undersöka vilken del AI skulle kunna ha. Visionen var att i ett senare skede kunna utveckla en helt automatiserad metod för att simulera ljudmiljön för vilken plats som helst i en stadsmiljö och på så sätt bidra till planeringen av stadsbygge. Projektet utgick från följande frågeställningar:

1. Hur kan auraliseringsmetoderna från *LISTEN* appliceras på en riktig plats för att simulera ljudsyntes i stadsmiljö?
2. Kan AI nyttjas för att ta fram ljudsyntes av samma plats?
3. Hur väl överensstämmer dessa modelleringar (auralisering och AI-modell) med verkligheten?

Auraliseringsmetoden syftade till att vara en förenklad modell och inkluderade därmed inte vindsamt temperaturpåverkan, vinklade ytor, ojämn mark, träd, buskar eller andra objekt som anses ha liten påverkan på den akustiska miljön. Mätningarna gjordes i enlighet med relevanta riktlinjer och praxis, enligt NORDTEST metoden[8], med undantag för avvikelser baserat på ett tydligt syfte eller omständigheter. Modellen bearbetar flerfilig trafik via hybridmetodiska auraliseringar som tillämpas på grund av projektets begränsade utsträckning.

2 Teori

I projektet har ljud modellerats från källor med signal i stabilt tillstånd som representeras av intensitet i dimensionerna av oktavband och direktivet. Oktavbanden delar upp ljudet i olika frekvenser och för varje band beskrivs även direktiviteten, hur ljudet är fördelat i olika riktningar. Auralisering är en process som syftar på att använda dessa källor för att simulera en ljudsignal som är representativ för det upplevda ljudet på en specifik plats. För att förklara metoden mer ingående förklaras konceptet samt dess grundläggande byggstenar i nästföljande avsnitt.

2.1 Auralisering

Den auraliseringsprocess som beskrivs i *LISTEN*-projektet börjar med extraktionen av ett käll-ljud från en inspelad fordonspassage [7]. Detta utfördes genom att i flera steg successivt ta bort akustiska egenskaper från ljudinspelningen. Dessa egenskaper är dämpningen i luften, markeffekten, direktiviteten, dopplereffekten, den sfäriska spridningen (avståndsförlust) och slutligen hur observatörens egen geometri påverkar hur ljudet når öronen. När käll-ljudet har extraherats kan dessa egenskaper sedan digitalt modelleras för en ny ljudfil med önskade avstånd samt markunderlag, vilket ger möjligheten att simulera ljudet hos ett fordon i en annan miljö med andra faktorer.

2.1.1 Dämpning från propagering

Ljudvågor som propagerar i medium kommer att förlora energi genom spridning och termiska förluster. Akustisk dämpning är ett mått på just denna energiförlust och kan för ljud vid en viss frekvens i utomhusmiljöer modelleras utifrån frekvensen, luftens temperatur, luftfuktighet, och lufttryck utifrån en ISO-standard [9].

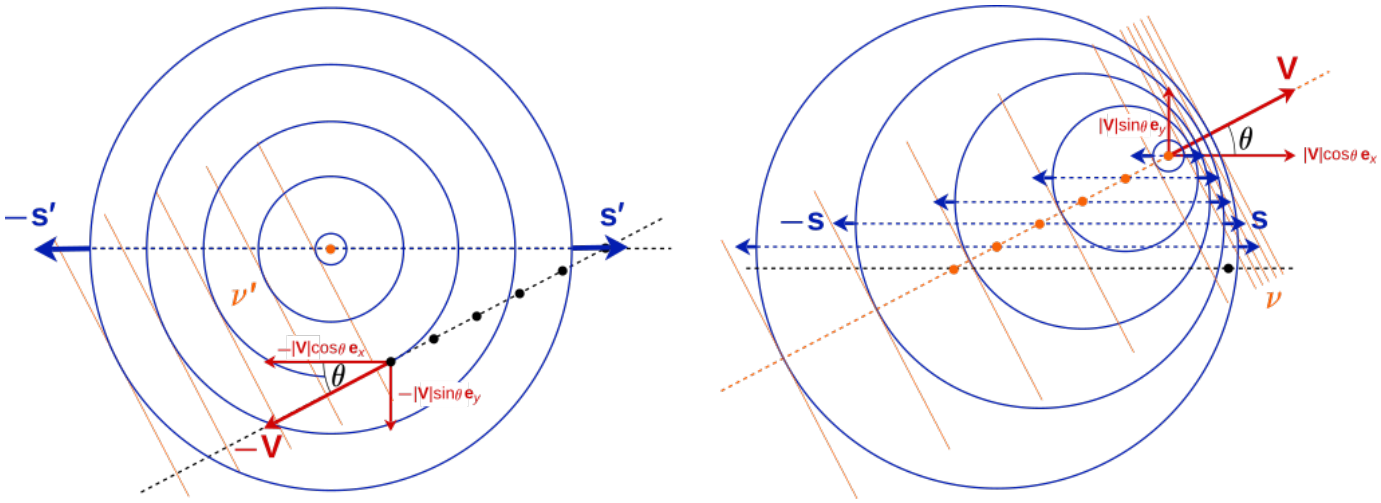
Utöver energi-förluster i mediumet självt kommer också, då ljudvågor sprids sfäriskt, intensiteten (energi per tid och area) att minska med radien [10, Kap. 5.3]. För en fast rymdvinkel är ytarean proportionerlig mot radien i kvadrat och således får vi intensitetförhållandet för sfärisk spridning $I \propto I_0 \frac{1}{r^2}$.

2.1.2 Markeffekt

För ljud som propagerar nära marken kommer reflektioner att påverka vågen hos mottagaren [11]. Detta kan modelleras med en plan markmodell där den primära vågen interfererar med den markreflekterade vågen. Storleken av denna effekt karakteriseras av ljudets våglängd och markens flödesresistivitet. Marktyper kan delas in i olika kategorier av flödesresistivitet σ beroende på hur kompakt marken är. Mjuk mark som exempelvis gräs och har låg flödesresistivitet $\sigma = 80$ kNsm^{-4} och hård mark som asfalt hög $\sigma = 200000$ kNsm^{-4} [11, Kap. 6.4.6].

2.1.3 Dopplereffekt

Då en ljudkälla och en observatör rör sig i förhållande till varandra kommer den uppfattade frekvensen hos ljudet skilja sig beroende på denna relativa rörelse [12]. Detta fenomen kallas för dopplereffekten och beror på högre eller lägre mottagen ljudfrekvens beroende på hastighet och riktning. Ljud sprider sig sfäriskt från sitt ursprung, vilket resulterar i att en stationär ljudkälla sprider ljudvågor med samma våglängd samt frekvens åt alla håll. Om ljudkällan istället är i rörelse kommer varje ljudvåg spridas sfäriskt från en ny position i förhållande till den föregående. Detta resulterar i en fördelning av ljudvågor där vågorna i källans rörelseriktning hamnar tätare varandra och det motsatta i den motsatta riktningen. En illustration av detta visas i figur 1.



Figur 1: Illustration av spridningen av vågor beroende på om ljudkällan är stationär eller inte. Till vänster illustreras den sfäriska spridningen, markerat som blå cirklar, från en stationär ljudkälla, den röda punkten, och till höger spridningen från en ljudkälla i rörelse vilken rör sig i riktningen markerad som V. Bildkälla: Maschen (CC0 1.0) [13].

En tätare packning av ljudvågor resulterar i en högre frekvens och en glesare packning i en lägre frekvens, vilket betyder att det uppfattade ljudet från en ljudkälla i rörelse kommer vara olika beroende på hur källan rör sig i förhållande till observatören. För en stationär observatör och

en rörlig källa som producerar ljud med frekvensen f och rör sig med en hastighet v relativt observatören kommer den upplevda frekvensen vara

$$f' = f\left(\frac{u}{u \mp v}\right) \tag{1}$$

där u är ljudhastigheten för mediumet. Tecknet för $\mp$ i nämnaren beror på källans färdriktning relativt observatören och är negativt för en källa som rör sig mot observatören.

2.2 Analys av ljud-data

Vid inspelning av ljud med en mikrofon mäts variationer i akustisk energi med hög tidsupplösning [14]. I ljudfilen sparas signalens amplitud i diskreta sampel, uttryckt med en viss sampelstorlek, b, normalt 16 bitar, alltså 2^{16} möjliga värden för amplituden [15]. Över ljudklippet är samplingsfrekvensen, f_s, antalet sampel per sekund (Hz), konstant och en vanlig frekvens för digitalt ljud är 44100 Hz. Samplingsfrekvensen bör enligt Nyquist teorem ligga på mer än dubbelt av den högsta frekvensen av signalen, den så kallade Nyquistfrekvensen, för att erhålla ett tillfredsställande resultat [16]. En ljudfil kan också ha flera signaler lagrade i ett antal kanaler, c. I detta projekt användes ljud i en kanal (mono) och två kanaler (stereo). Med detta kan filstorleken för okomprimerat ljud beräknas som

$$\text{Filstorlek (Bytes)} = \frac{f_s \cdot b \cdot c \cdot t}{8} + H \tag{2}$$

där H är storleken på formatspecifikationer och annan metadata som sparas i filen utöver den faktiska ljuddata och är ofta väldigt liten i jämförelse med totala filstorleken.

Vid analys av ljud-data är det dock ofta väldigt meningsfullt att analysera periodiska mönster i erhållen data och med vilka frekvenser dessa mönster återfinns. För att ta data som existerar i tidsdimensionen till en dimension uttryckt av frekvenser kan fouriertransform användas [17]. För numeriska beräkningar används oftast någon form av *Fast Fourier transform* (FFT) algoritm. Något som kan vara viktigt vid numerisk fouriertransform är hur signalen uppför sig i kanterna av data-segmentet [17]. Vid FFT antas att data är perfekt periodisk över segmentet vilket så klart aldrig kommer att ske vid faktiska mätningar. För att lösa detta multipliceras data-segmentet med en fönsterfunktion (exempelvis Hann-fönster används i detta projekt) som går mot 0 vid intervallets ändar och således blir segmentet alltid kontinuerligt i start och slut.

2.3 Statistisk analys

För att kunna analysera samt utvärdera resultaten från lyssningsförsök med frivilliga deltagare behöver vissa statistiskt analytiska modeller tas i åtanke. De modeller av störst intresse för detta projekt var Friedman-testet, Kendalls konkordanskoefficient, Wilcoxon-testet och konfidensintervall för binomial data.

2.3.1 Icke-parametriska rangtester

Friedman-testet är ett icke-parametriskt test vilket syftar på att jämföra tre eller fler beroende grupper [18, Kap. 7.1]. Metoden bygger på att samma enhet observeras under olika förhållanden, kallade betingelser. En betingelse avser ett specifikt villkor som den beroende enheten utsätts för. Upprepade observationer inom samma enhet bildar ett så kallat block. Inom varje block rangordnas sedan varje observation och ersätts med ett rangvärde R_{ij} där i indexerar blocket och j betingelsen. Detta skulle exempelvis kunna illustreras som att en individ (ett block) ger ett betyg på hur god en viss typ av glass är (en betingelse). Därefter ersätts dessa betyg med ett rangvärde R_{ij}, baserat på betyget en glass får i förhållande till de andra. Teststatistiken S definieras som

$$S = \frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1), \tag{3}$$

där n är antalet block, k antalet betingelser och R_j summan av rangvärdena R_{ij} över alla block för betingelse j [18, Kap. 7.1]. Teststatistiken S används sedan för att avgöra ifall en nollhypotes, vilken antar att inga systematiska skillnader mellan betingelserna finns, är berättigad [18, Kap. 7.1]. Genom att jämföra S med en chitvåfördelning (χ^2), från tabellerad data [19], med antalet frihetsgrader $k-1$ kan nollhypotesen förkastas då $S \geq \chi^2_{k-1,\alpha}$.

För att avgöra hur starkt rangordningarna över betingelserna skiljer sig åt eller liknar varandra kan sedan Kendalls konkordanskoefficient (Kendall's W) användas[18, Kap. 8.1]. Till skillnad från Friedman-testet som berättar ifall det finns en signifikant statistisk skillnad mellan betingelser sätter Kendall's W ett mått på styrkan i denna skillnad. Koefficienten W definieras som

$$W = \frac{12S}{n^2(k^3 - k)}, \tag{4}$$

där n är antalet block, k antalet betingelser och S, inte att förväxla med teststatistiken från Friedman-testet, summan av kvadrerade avvikelser från medelrangen, $S = \sum_{j=1}^{k}(R_j - \bar{R})^2$ [20, 21]. W kan anta värden mellan noll och ett, där $W = 0$ innebär ingen överensstämmelse mellan rangordningar och $W = 1$ innebär en fullständig överenskommelse mellan rangordningar [18, Kap. 7.1].

I slutändan analyserar varken Friedman-testet eller Kendall's W skillnader mellan två specifika betingelser, vi kan alltså inte titta på hur två specifika betingelser förhåller sig till varandra. Därav introduceras Wilcoxon-testet, även kallat *Wilcoxon signed-rank test*. Wilcoxon-testet utgår ifrån skillnader mellan observationer hos två betingelser i varje block [18, Kap. 3.1]. För varje block beräknas därav $D_i = X_i - Y_i$, där X_i är observationen hos en betingelse inom block i och Y_i observationen för en annan betingelse i samma block.

Absolutbeloppen $|D_i|$ rangordnas sedan över samtliga block, där det minsta värdet erhåller rang 1, det näst minsta rang 2 och så vidare [18, Kap. 3.1]. I det fall två eller flera $|D_i|$ har samma värde tilldelas dessa ett genomsnittligt rangvärde. I det fallet $|D_i|$ istället är noll för ett block exkluderas detta. Efter att rangordningen fastställts introduceras tecknet från den ursprungliga skillnaden tillbaka och rangvärdena sorteras in i grupper av positiva T^+ eller negativa T^- observationer [18, Kap. 3.1]. Värdena i dessa skilda grupper summeras sedan som

$$T^+ = \sum R_i^+ \tag{5}$$

$$T^- = \sum R_i^- \tag{6}$$

där R_i^+ är de positiva rangvärdena och R_i^- de negativa. Slutligen definieras teststatistiken som Q

$$Q = \min(T^+, T^-), \tag{7}$$

vilket alltså betyder att Q antar värdet på den summa med minst värde. Om nollhypotesen varit att ingen signifikant skillnad finns mellan de två betingelserna kan denna förkastas ifall värdet på Q är mindre eller lika med ett kritiskt värde från Wilcoxons fördelning[22].

2.3.2 Konfidensintervall för binomial data

Konfidensintervall används inom statistik för att uppskatta inom vilket intervall ett okänt populationsvärde sannolikt befinner sig baserat på observerad stickprovsdata [23]. Här syftar populationsvärdet på det verkliga värdet av en statistik undersökning över en större population än vad stickprovsdatan representerar. Ett konfidensintervall ger därav en skattning på vart detta värde ligger samt osäkerheten i denna skattning. I detta arbete användes konfidensintervall på en 95%-nivå. Ett sådant intervall innebär att ifall en motsvarande undersökning utförts ett stort antal gånger med nya stickprov skulle ungefär 95% av de beräknade intervallen förväntas behålla populationsvärdet [23].

Då binomial data hanteras, data där enbart två olika utfall finns, kan Wald metoden användas [24]. Om X är antalet observationer av ett visst utfall och n det totala antalet observationer kan den observerade andelen uttryckas som

$$\hat{p} = \frac{X}{n}, \tag{8}$$

där $\hat{p}$ representerar stickprovets uppskattning av den sanna andelen observationer i populationen. För tillräckligt stora stickprov, då $n \cdot \min(\hat{p}, 1-\hat{p}) > 10$, kan $\hat{p}$ approximeras som normalfördelad och konfidensintervallet uppskattas som

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \tag{9}$$

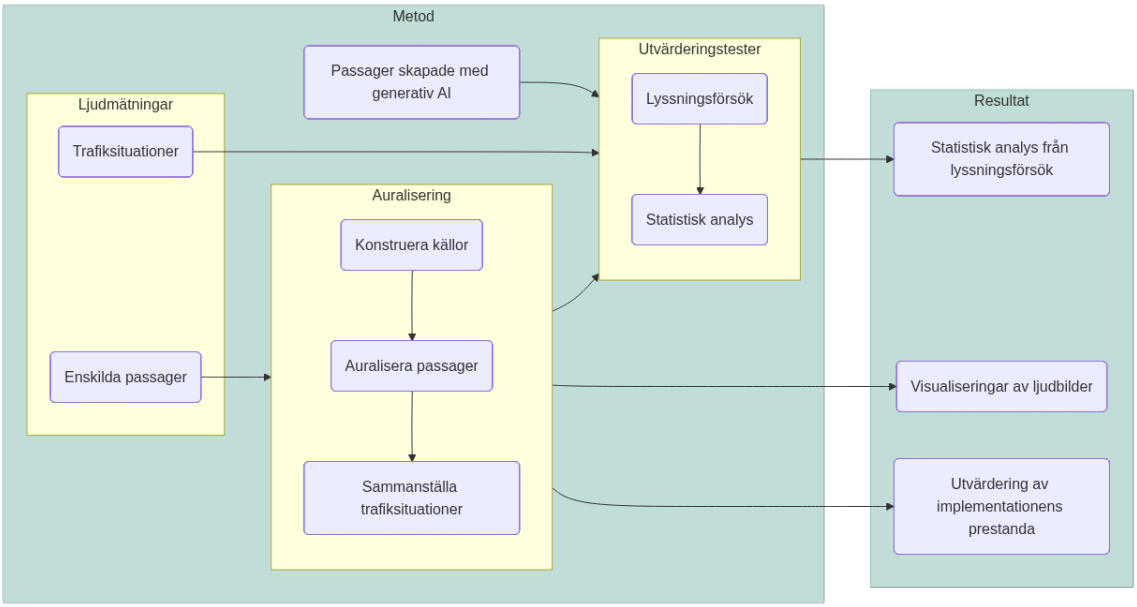
där $z_{\alpha/2}$ percentilen från standardnormalfördelningen motsvarande en vald signifikansnivå α [24]. För ett konfidensintervall på 95%-nivå används då $z_{0,025} = 1{,}96$ [19].

3 Metod

Arbetsprocessen för genomförandet av kandidatarbetet har inkluderat ett flertal moment med syfte till att undersöka möjligheterna att framställa ljudsynteser av verkliga stadsmiljöer med utgångspunkt i de modeller som användes i *LISTEN*. Detta genomfördes via förstudier, datainsamling,

analyser och utvärderingar. Projektet inleddes med en förstudie av befintlig litteratur inom området främst fokuserat på det som presenteras i *LISTEN* [7], men även andra vetenskapliga artiklar, rapporter och liknande studier samt examensarbeten bearbetades för att få en överblick över tidigare forskning, metodik och kunskap inom området. Litteraturstudien genomfördes även för att skapa en teoretisk grund samt för att identifiera de begränsningar, utvecklingsområden och kunskapsluckor som motiverat projektets syfte.

Metoden för projektet presenteras i fyra delar: ljudinspelningar, auralisering, tillgängliga AI-modeller och utvärderingstester. För att kunna presentera resultat och svara på frågeställningarna har alla delar genomförts och i rätt ordning då de bygger på varandra, se figur 2. Varje del i metoden har inte för sig självt ett relevant resultat att presentera men är en del av processen för att kunna ta fram de delar som senare presenteras i resultatet, kap. 4.

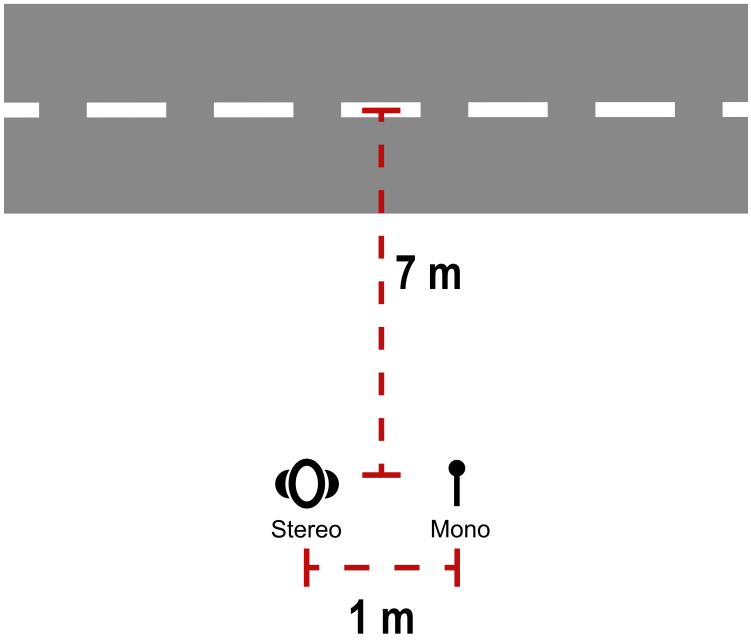


Figur 2: Flödesschema över de olika delarna av projektet som presenteras i metoden, hur de bygger på varandra och bidrar till de resultat som presenteras.

3.1 Ljudinspelningar

Ljudinspelningar gjordes med två olika syften, dels för att användas som källor för auralisering och dels för verifiering av auraliseringar av mer komplexa trafik-situationer. Mätningar genomfördes vid valda platser för att representera de olika miljö- och trafikförhållanden som projektet syftade till att auralisera och simulera.

Mätningarna för käll-ljud gjordes med ett avstånd på 7 meter från mitten av vägen som avsedde mätas. Höjden på mono-mikrofonen var 1,20 meter och höjden på stereo-mikrofonen placerad på huvudstativ var 1,40 meter. Mikrofonerna placerades med ett avstånd på ca 1 meter ifrån varandra. Avstånd och placering varierade över mätningarna beroende på avsikt. Mätningarna vid den sista mätplatsen bestod av en inspelning avsedd att samla in data för en mer komplex trafiksituation och genomfördes därmed inte strikt enligt dessa riktlinjer i samma utsträckning. Den generella uppställningen av mätutrustningen redovisas i figur 3.



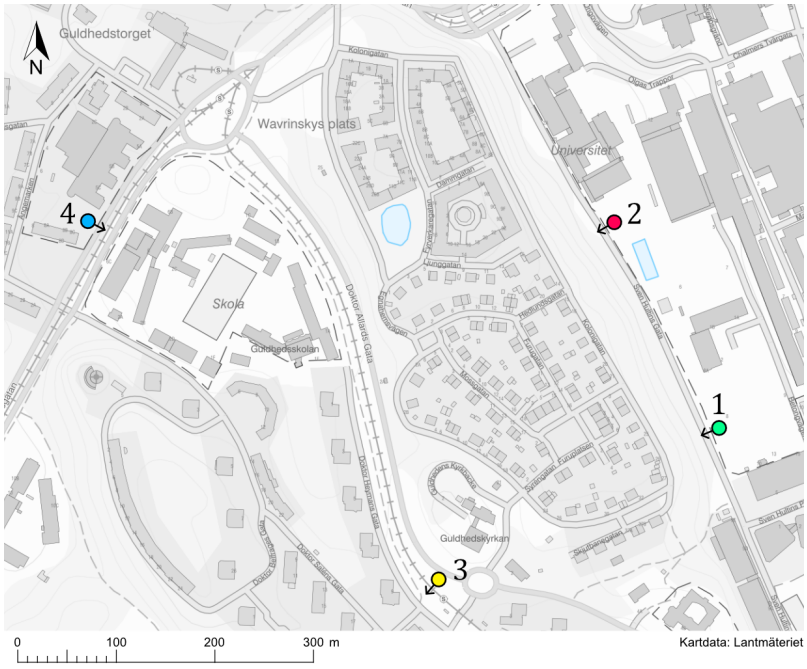
Figur 3: Illustrering av mikrofonpositionering i förhållande till vägbanan.

Efter avvägning utifrån första mättillfället gjordes diverse ändringar för att underlätta insamlingen av data. Vid första mättillfället inspelades enstaka passager med en tid på ca 10-20 sekunder. Andra mättillfället ändrades denna metod till att spela in under en längre period, ca 3-6 minuter i samband med videoinspelning för att i efterhand kunna göra en mer noggrann bedömning av fordonsmodeller, hastighet, avvikelser etc. Dessa inspelningar sammanställdes därefter med inspelningsprogram som *Audacity* och *Logic Pro X*. De sammanställda ljudsignalerna användes sedan som referensmaterial och underlag till modelleringen och lyssningsförsöken. Mätningarna förutsattes att genomföras enligt vedertagna riktlinjer i den utsträckning det varit möjligt för arbetets tidsram och tillgångar.

Mätutrustningen som användes för att genomföra inspelningar inkluderade GRAS 146AE 1/2” CCP frifältsmikrofon för monoinspelningar och HEADacoustics SQobold 3302 för binaurala stereoinspelningar. Båda mikrofoner användes med tillhörande vindskydd för att minimera påverkan av vindstörningar i ljudsignalerna.

3.1.1 Inspelningstillfällen

Mätningar utfördes på fyra platser och tillfällen. Se översiktskartan över mätplatser, figur 4, och tabell 1 för mer detaljerad information om platserna.



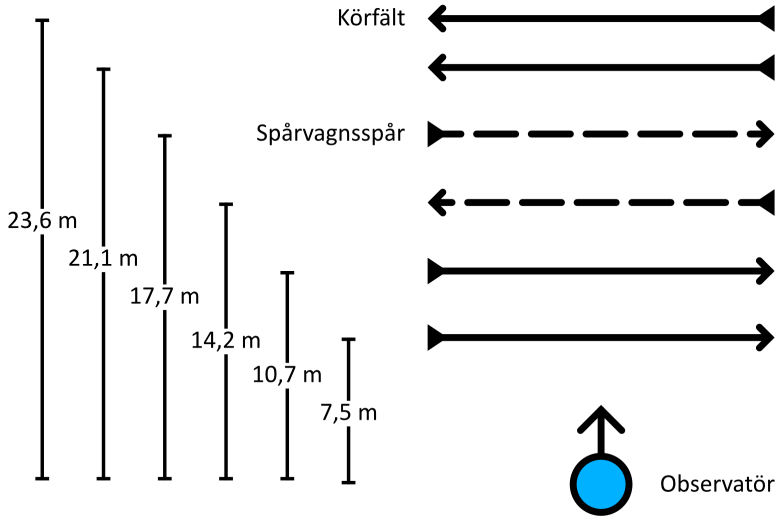
Figur 4: Översiktskarta över mätplatser. Plats 1 och 2 gjordes mätningar av källjud vid olika tillfällen men båda mot Sven Hultins gata vid Chalmers campus. Vid plats 3 gjordes mätningar på spårvagnar mot spåren nordväst om stationen Doktor Fries Torg. Vid plats 4 gjordes mätningar av mer komplexa trafiksituationer. Pilen vid punkterna pekar åt det håll mikrofonerna riktats.

Tabell 1: Väderförutsättningar för ljudinspelningar vid de olika mättillfällena.

Plats	1	2	3	4
Datum	2026-02-26	2026-03-02	2026-04-01	2026-04-09
Tid	10:00:00	13:00:00	10:00:00	10:00:00
Temperatur [°C]	4,9	5,7	5,0	10,5
Vindhastighet [m/s]	5,1	1,9	3,1	3,1
Luftfuktighet [%RH]	95	95	99	56

Mätningar i syfte att användas som käll-ljud gjordes mot Sven Hultins gata, plats 1 & 2 i översiktskartan, se figur 4. Trafikintensiteten längs denna gata var låg vilket gjorde den bra lämpad för mätning av enskilda fordon. Längs vägen passerade främst fordon i personbilsstorlek med hastigheter runt 30 km/h. Det första utav dessa mätningarna, plats 1, kan anses som en testmätning, då förhållanden var något sämre med måttlig vind och blött underlag. Mätningar från plats 1 användes inte i några källor för auralisering då avvikelser i ljudmiljön förekom som sågning av träd samt flera förbipasserande gång- och cykeltrafikanter. Senare mätningar, vid plats 2, hade bättre förutsättningar men även här en något fuktig vägbana.

Mot spårvagnsspåret ca 30 m nordväst om stationen Doktor Fries torg, plats 3, gjordes mätningar på spårvagnar. Mikrofonerna placerades på samma sätt som i mätningar av bil-passager, se figur 3, och mätningar gjordes på båda spåren. Eftersom att mätningen utfördes i närhet till en spårvagnsstation uppmättes åt ena hållet spårvagnar som bromsade in något och åt andra hållet vagnar som accelererade.



Figur 5: Modell över körfält vid mätplats 4.

Vid plats 4 gjordes längre mätningar av mer komplexa trafiksituationer i syfte att användas som referens-fall för att skatta hur väl modellerna representerar verkligheten. Förutsättningarna för mätning var mycket goda med torrt väglag och låga vindhastigheter med avståndet 7,5 meter till mitten av det närmsta körfältet, se figur 5. Detta är ett större avstånd än angivna riktlinjer eftersom mätningen hade för avsikt att spela in en mer övergripande ljudbild av en komplex trafiksituation som inkluderar spårvagn, buss, tung trafik och personbilar.

3.2 Auralisering

Auraliseringsprocessen från källinspelningar till modellering av en trafiksituation utfördes i flera steg. Först gjordes mätningarna av enskilda fordonspassager om till källsignaler i stabilt tillstånd. Därefter användes dessa källor för att modellera passager på olika avstånd, hastigheter och fordonstyper. Slutligen kombinerades flera passager för att skapa mer komplexa trafiksituationer.

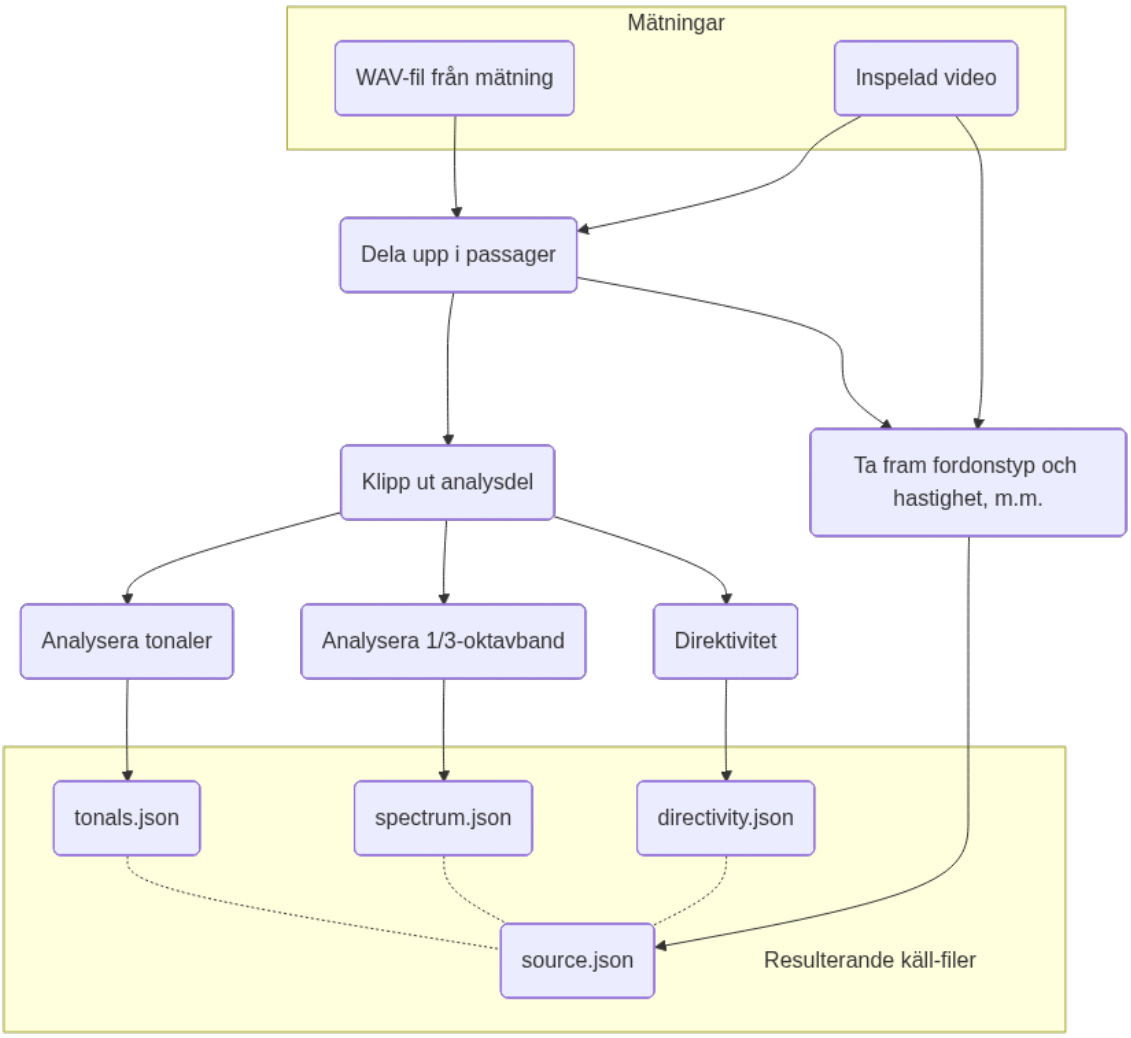
Implementering skedde till stor del i Python. Auralisering av passager baserades metodmässigt helt på implementeringskod i MATLAB från *LISTEN*. Denna skrevs om till Python för att utifrån projektgruppens bakgrundskunskaper underlätta vidare utveckling och mer frihet att utforma källornas format. Koden finns att tillgå via Git, se appendix B. *Github Copilot* användes som verktyg vid implementeringen, i störst utsträckning och i *agent mode* vid konvertering från MATLAB till Python men även som enklare stöd, och för att kontinuerligt skriva dokumentation under de vidareutvecklingar som gjordes.

3.2.1 Käll-ljud

De längre inspelningarna i syfte att skapa käll-ljud klipptes ner till flera kortare ljud-klipp för varje fordonspassage. Genereringen av källsteg genomfördes i flera steg från inspelning till JSON-filer, se figur 6. För käll-ljud användes ljudet från den riktade mikrofonen, så alltså bara en kanal. Från den video som spelades in beräknades fordonens hastighet utifrån uppmätt längdreferens och utifrån registreringsnummer kategoriserades även fordonsmodeller. Fordonsinformation om de källljud som faktiskt användes finns i appendix C. För att forma de tids-statiska källorna används bara en liten del (0,5 sekunder) av inspelningen, precis efter det att fordonet passerat rakt framför mikrofonen. Denna korta ljudsignal delas upp ytterligare i korta segment vars frekvensspektra tas fram med FFT. De högupplösta spektra integreras över varje band för att forma frekvensdata som tredjedels oktavband. Bandvisa tidsmedelvärden över samtliga segment formar tidsstatiska källfrekvenser uttryckt i dessa band. I källornas spektrum sparas också identifierade ordningstoner i form av sinus-toner och FFT-puls som används för att återskapa en harmonisk struktur som kan fördelas över passagen. Det kompenserades sedan för akustiska egenskaper under själva mätningen. Propagerings-förluster, sfärisk spridning och markeffekt beräknades bort så att källan representerar ljudet precis vid fordonet (1 m ifrån källan som referens för intensitet vid sfärisk spridning).

Varje källa hade också en varierande förstärkning beroende på riktning, så kallad direktivitet. På grund av den begränsade mätuppställningen kunde ingen experimentellt motiverad direktivitetsprofil tas fram så istället användes en profil från *LISTEN* med mycket nära omnidirektionell fördelning.

Rent praktiskt sparades källdata i json-filer uppdelade i *source*, *spectrum* och *directivity*. *Source* innehöll generell data om källan som exempelvis fordonstyp, hastighet, referensavstånd, samt länkar till json-filerna för spektrum och direktivet. *Directivity* innehöll källans direktiva förstärkning antingen som ett polynom eller som en lista med vinkel och förstärkning. För alla källor användes samma direktivitet för samtliga band även om det var möjligt att ha olika. *Spectrum* innehöll källans faktiska spektrum som tredjedels oktavband, och tonala komponenter uttryckt FFT-baserat och som sinusfunktioner.

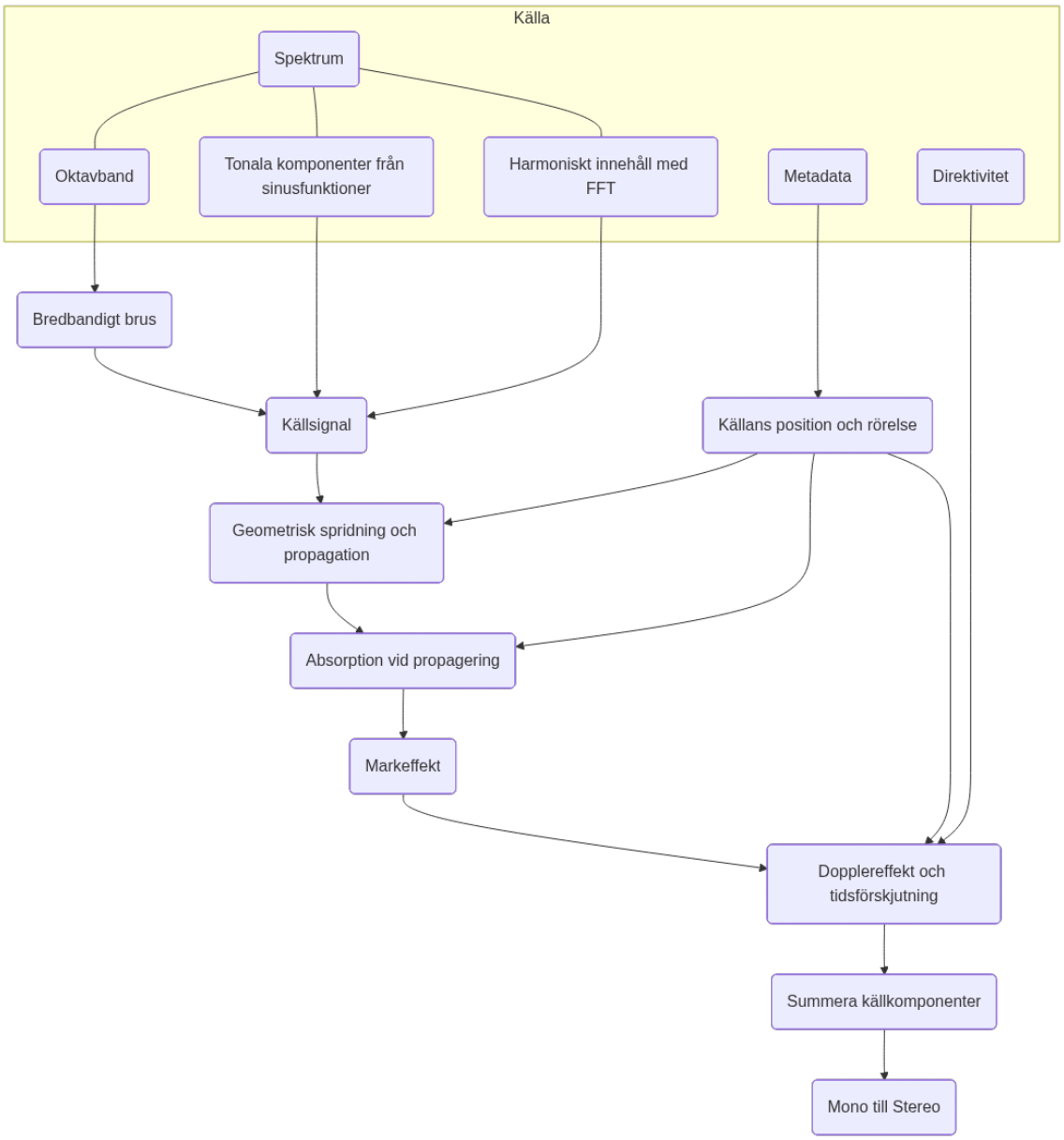


Figur 6: Flödesschema över processen att skapa källor från passage-mätningar.

3.2.2 Enskilda passager

Utifrån de framtagna källorna kunde sedan de auraliseringssteg som beskrivs i kap. 2.1 användas för att modellera fordonspassager på olika avstånd. Utifrån källans frekvensband genererades en källsignal genom en kombination av bredbandigt brus i oktavband, tonala komponenter från sinusfunktioner samt harmoniskt innehåll rekonstruerat med FFT-metoder. När källsignalen hade skapats modellerades ljudets propagation genom geometrisk spridning, luftabsorption och markeffekter. Därefter applicerades källans direktivitet och rörelse för att beskriva hur ljudutstrålningen varierar med vinkel och position över tid. Slutligen applicerades Dopplereffekt och tidsförskjutning

baserat på fordonets rörelse relativt mottagaren, varefter alla komponenter summerades till den slutliga auraliserade ljudsignalen, se figur 7. Här exporterades de auraliserade passagerna till wav-filer i mono. Dessa kunde sedan processas i ett av examinator J.Forssén (privat kommunikation, 20 april, 2026) givet MATLAB-program som genererade stereoljud utifrån källans position och rörelse, samt en modell av hur en mänsklig observatör skulle uppleva ljudet.



Figur 7: Flödesschema över processen att skapa auraliserade ljudsignaler ifrån framtagna källor.

3.2.3 Mer komplexa trafiksituationer

Den använda auraliseringsmetoden möjliggjorde simulering av enskilda fordonspassager, men var begränsad i möjligheten att skapa mer komplexa situationer och ljudmiljöer med flera samtidiga passager. Då detta undersökningsområde fortfarande var av intresse för projektet, samt relevant för framtida utvecklande av auraliseringsmetodik, kompletterades den automatiserade processen med manuell sammanställning. Därav testades två skilda metoder för att sammanställa de komplexa trafiksituationerna.

Den första metoden, metod 1, utfördes genom att sammanföringen av ljudfiler använde sig av de stereosimuleringar som gjordes via auralisering enligt skriven MATLAB-kod. För att sammanställa flera ljudkällor för att skapa mer komplexa trafiksituationer utnyttjades programmet *Logic Pro X* genom att överlappa och sammanföra flera enskilt auraliserade ljudspår.

I den andra hybridmetoden, metod 2, sammanställdes flera ljudkällor till en mer komplex trafiksituation på samma sätt som innan. Utöver detta så binauralt panorerades de sammansatta auraliserade mono-ljudfilerna i horisontalplanet utifrån deras antagna position i förhållande till lyssnaren. Den binaurala panoreringen hade för avsikt att efterlikna realistiska trafikscenarier och bör betraktas som en perceptuell approximation i syfte att uppfylla och möjliggöra explorativ analys, snarare än en fullständigt utvecklad metodik eller fysiskt exakt simulering.

Båda tillvägagångssätten innehöll en ljudfil med en spårvagnsauralisering. Då auraliseringen av spårvagnen inte lyckades genomföras på samma sätt som bilpassagerna bör det noteras att den inte behandlades i samma utsträckning och blev endast processad genom normalisering av ljudnivån. Alltså är spårvagnsljudet i filerna inte fullständigt auraliserat.

För att skapa jämförande material inför lyssningstesten valdes tre trafiksituationer från inspelat material med en varaktighet på ca 5-10 sekunder långa, där flera passager av olika fordonstyper förekom. Med utnyttjande av movie-funktionen i *Logic Pro X*, som möjliggör synkning av ljud och bild, så genomfördes försök till att återskapa dessa trafiksituationer. Detta gjordes via utvalda auraliseringsklipp som motsvarade de passager som identifierades på ljudinspelningen samt de fordon som observerades i videon. Utöver dessa rekonstruktionsförsök som syftade till direkt jämförelse och utvärdering, skapades även tre syntetiska trafiksituationer utan koppling till inspelat material på samma vis. Vidare inkluderades även tre ljudklipp tagna ifrån inspelningen som inte hade koppling till något av de sammanställda auraliseringarna för att erhålla en bredare bedömningsgrund av auraliseringskvalitén.

3.3 Evaluering av tillgängliga AI-modeller

Under arbetet genomfördes även prövningar och försök till att generera simulerade ljudfiler via offentligt tillgängliga modeller för generativ AI. De generativa AI-modeller med förmågan att skapa ljudfiler som prövades i projektet inkluderade Suno, Optimizer AI samt ElevenLabs. Via försök att prompta fram ljudsimuleringar för trafikflöde och passager av fordon lyckades endast Optimizer AI samt ElevenLabs att få fram önskvärda resultat. Suno tog endast fram musikfiler oavsett prompt och kunde inte generera en ljudfil i efterfrågat format. Optimizer AI samt ElevenLabs är program

avsett för att skapa ljud. Optimizer AI grundades som ett företag med avsikt att forska och utveckla AI inom ljuddesign främst avsett för industrier som film samt dator- och TV-spel [25]. ElevenLabs grundades för att generera mänskliga röster via AI men har sedan dess expanderat till att även generera ljudeffekter, musik, video och mer[26].

Även om Optimizer AI hade god förmåga att genera efterfrågade ljudfiler utifrån prompts var den begränsad i sin förmåga att skapa mer komplexa ljudmiljöer. Exempelvis kunde modellen inte generera tillfredsställande ljudfiler utefter prompts som begärde mer tung och frekvent trafik eller flera ljud samtidigt. Enklare och mer dramatiska ljudeffekter funkade bättre att generera. Som experiment att testa modellens förmågor gjordes försök till att prompta fram en bilkrasch där en bil sladdar och sedan krockar. Detta lyckades med godtyckligt resultat vilket visar på att modellen har förmågor att skapa ljudeffekter och scenarion med relativt trovärdigt resultat, men har ännu inte förutsättningar att generera mer komplexa ljudmiljöer. Optimizer AI har även en funktion som tillåter skapande av variationer av ett uppladdat ljud. Denna funktionen utvärderades också med ett klipp från en bilpassage och resultatet blev relativt tillfredsställande utifrån det uppladdade klippet. Den främst märkbara skillnaden var avsaknaden av de små detaljerna i ljudbilden så som knaster från gruskorn under däcken och liknande.

I slutändan valdes dock ElevenLabs att användas, då Optimizer AI inte tillät nedladdningar av skapade ljudfiler utan betalning. Plattformen erbjuder olika typer av AI-genererade ljud, men påstår sig kunna skapa realistiska ljudfiler av vilket ljud som helst genom sitt "*Sound Effects*" verktyg[27]. Olika prompts testades och exempel på ett par av de som användes presenteras nedan.

Prompt exempel 1

The sound of heavy traffic flow from a receiver standing besides the road.

Prompt exempel 2

City ambience with nearby car passing by, one pass, rather slow.

Prompt exempel 3

The sound of heavy traffic flow from a receiver standing besides the road in an urban landscape. Some birds tweeting should be heard in the background.

I slutändan noterades däremot att ElevenLabs inte var kapabel till att skapa de exakta situationer som var önskade. Komplexa trafiksituationer kunde inte genereras och alla prompts som användes gav ungefär samma resultat, där resultaten alltid var passagen av ett fordon. Däremot skiljde sig dessa enskilda passager relativt mycket åt och bedömdes vara tillräckliga för att bidra till projektet.

3.4 Utvärderingstester

Framtagna ljudklipp utvärderades i slutskedet genom lyssningsförsök med frivilliga deltagare. Deltagarna fick lyssna på både inspelade och syntetiskt genererade ljudexempel och ombeds bedöma bland annat den upplevda realismen hos ljudklippen. Resultaten analyserades sedan för att bedöma kvaliteten av de olika metodikerna och adressera deras individuella trovärdighet samt att undersöka eventuella problem och vidare utveckling.

3.4.1 Lyssningstest med deltagare

I syfte att validera ljudfilerna som skapats genom samtliga metodiker utfördes ett lyssningstest med hjälp av frivilliga deltagare. Testet syftade på att pröva den upplevda realismen av ljuden, ifall de uppfattades vara inspelade eller genererade och vilken av de skapade ljuden som upplevts mest trovärdigt återspegla den verkliga inspelade ljudmiljön.

Testet utfördes individuellt av varje deltagare genom programvaran Artemis Suite och med tillgång till samma typ av hörlurar samt dator. Innehållet i testet skiljde sig inte åt mellan deltagarna och bestod av korta ljudklipp, ca 5-8 sekunder, som spelades upp med en tillhörande fråga, där ordningen på ljudklippen på varje sida kom i en slumpmässig ordning för varje deltagare. Varje deltest bestod av flera ljudklipp från varje ljudkälla, förutom i den sista delen då de AI-genererade verktygen inte var kapabla till att simulera mer komplexa ljudmiljöer. För varje del i testet hade deltagarna tillgång till en riktigt inspelning av ett fordon som referens och även möjligheten att skriva ned kommentarer om varför de bedömt ett visst ljudklipp som de gjort.

Under testets gång utförde maximalt fyra individer testet samtidigt, med ljudskärmar mellan sig för att minska läckage mellan hörlurar. Sammanlagt deltog 15 individer, varav 13 män och 2 kvinnor, i olika åldrar över spannet 20-25 år. Av deltagarna hade enbart 2 individer utfört liknande tester tidigare, men enbart vid en till två tillfällen.

För att utvärdera den upplevda realismen hos ljudklippen fick deltagarna först lyssna på ett antal ljudklipp och bedöma realismen på dessa på en likertskala mellan 1-7. Ett betyg på 1 återspeglade en mycket orealistiskt bedömd upplevelse och ett betyg på 7 en mycket realistiskt bedömd upplevelse. Ljudklippen bestod under denna del enbart av simpla trafiksituationer, vilket exempelvis skulle kunna vara ljudet från passagen av ett fordon. Ljudklippen som spelades kom i slumpmässig ordning och deltagarna var ovetandes om ljudet skapats eller spelats in. Deltagarna fick totalt lyssna på tolv olika ljudklipp, varav fyra var inspelade, fyra var auraliserade och fyra var AI-genererade.

I den andra delen av testet prövades ifall deltagarna kunde särskilja på ifall ett ljudklipp var skapat eller inspelat. Deltagarna utförde då samma process som i den första delen, men fick lyssna på andra ljudklipp samt enbart besvara ifall de trodde klippet var skapat eller inspelat. I det fall deltagaren inte tydligt kunde särskilja ifall ljudet var syntetiskt eller ej fanns även alternativet “Vet ej” som svar. Totalt fick deltagarna lyssna på tolv olika ljudklipp, varav fyra var inspelade, fyra var auraliserade och fyra var AI-genererade.

Den sista delen bestod av en serie rankningar, med samma likertskala som i den första delen, av den upplevda realismen av auraliserade ljud samt inspelade ljud i mer komplexa trafiksituationer. Deltagarna fick först lyssna på tre trafiksituationer från båda metodikerna som simulerade slumpmässiga fall. Därefter fick de lyssna på 2 olika specifika fall, där de auraliserade ljuden hade försökt återskapa en så lik ljudmiljö till en referensinspelning som möjligt. I det första fallet lyssnade deltagarna på en inspelning samt ett auraliserat ljud, utan att veta vilken som var vilken, och fick sedan betygsätta realismen hos de två ljudklippen. I det andra fallet användes en annan inspelning som referens och de två auraliserade ljudklipp som skapats på olika sätt. Deltagarna fick därpå, som i det första fallet, betygsätta realismen hos de två auraliserade ljudklippen samt inspelningen utan att veta vilken som var vilken.

3.4.2 Analysmetod av lyssningstest

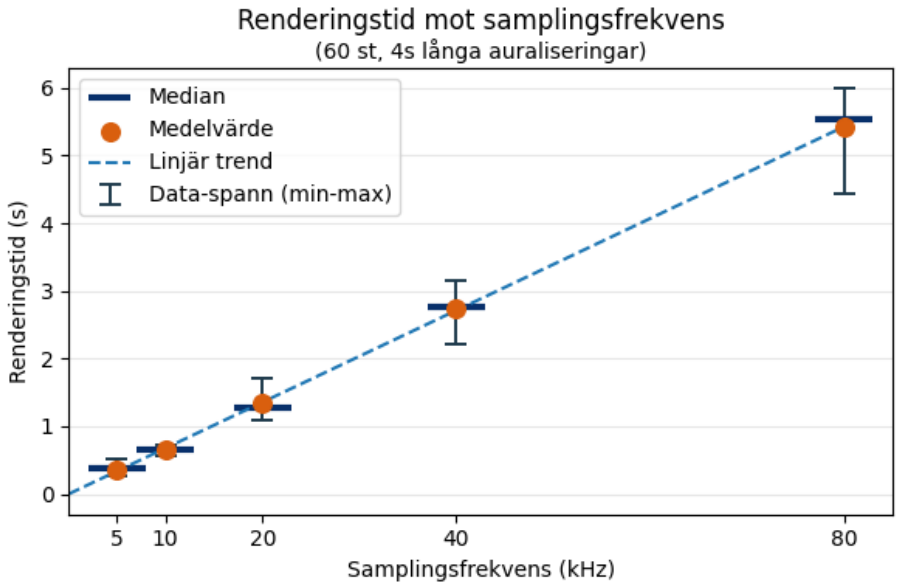
Då lyssningsförsöken utförts sammanställdes all data för att statistiskt analyseras i enlighet med de statistiska metoder som beskrivs i 2.3. För den första och sista delen av testet sammanställdes först medelbetyget av varje ljudkällas ljudklipp för varje person. Därefter applicerades ett Friedman-test för att avgöra ifall det fanns en signifikant statistisk skillnad mellan rangordningar av ljudklippen från de olika ljudkällorna, måttet på styrkan av den skillnad beräknades också enligt Kendalls konkordanskoefficient. I det fall en signifikant skillnad uppskattades enligt Friedman-testet utfördes ett individuellt Wilcoxon-test mellan varje ljudkälla för att se hur skillnaderna förhöll sig mellan dessa. För den andra delen av testet då enbart binomial data erhölls beräknades det procentuella måttet på hur ofta deltagarna korrekt kunde avgöra ifall ljudklippen ifrån en ljudkälla var skapade eller inspelade, samt ett konfidensintervall på 95%-nivå för varje ljudkälla. Kommentarer från deltagarna analyserades sedan för att urskilja förbättringsområden och styrkor hos de olika ljudkällorna.

4 Resultat

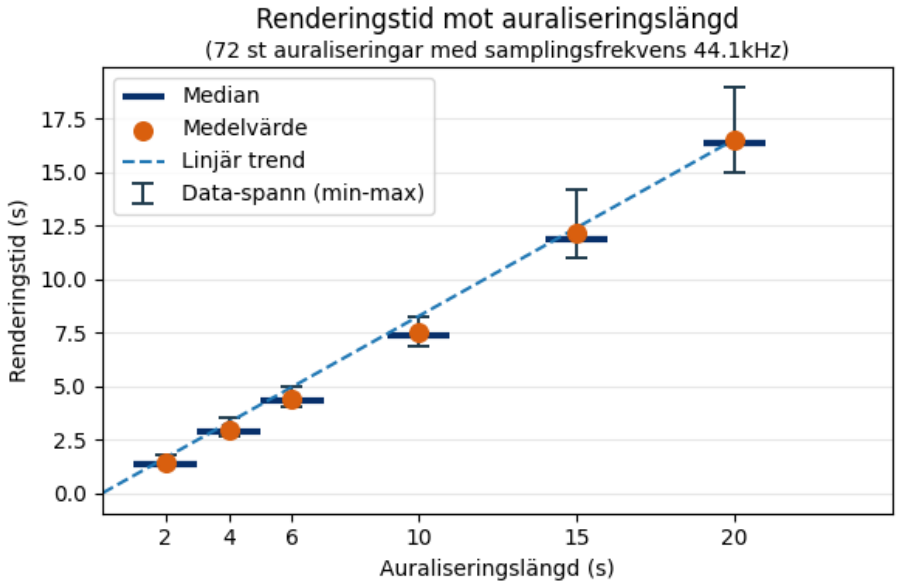
I följande avsnitt presenteras relevanta resultat från samtliga delar av projektet. Först presenteras resultaten för prestandan av auraliseringsprocessen och därefter grafiska visualiseringar av framställda ljudfiler. Slutligen redovisas de statistiskt analytiska resultaten för lyssningstestet, sorterat efter testets 3 delar.

4.1 Prestanda

Implementeringen av auralisering för enskilda passager som gjordes i Python kördes för olika källor, avstånd, samplingsfrekvenser och längd på ljudfil. I figur 8 och 9 utläses hur renderingstiden är linjärt proportionerlig mot både storleken på samplingsfrekvens och auraliseringens längd. Filstorleken för de genererade filerna i standardfallet med 5 sekunders långt ljudklipp, samplingsfrekvens 44 100 Hz, sampelstorlek 16 bitar och 2 kanaler blev ca 860 kB vilket var marginellt lägre än vad som ges med ekv. 2.



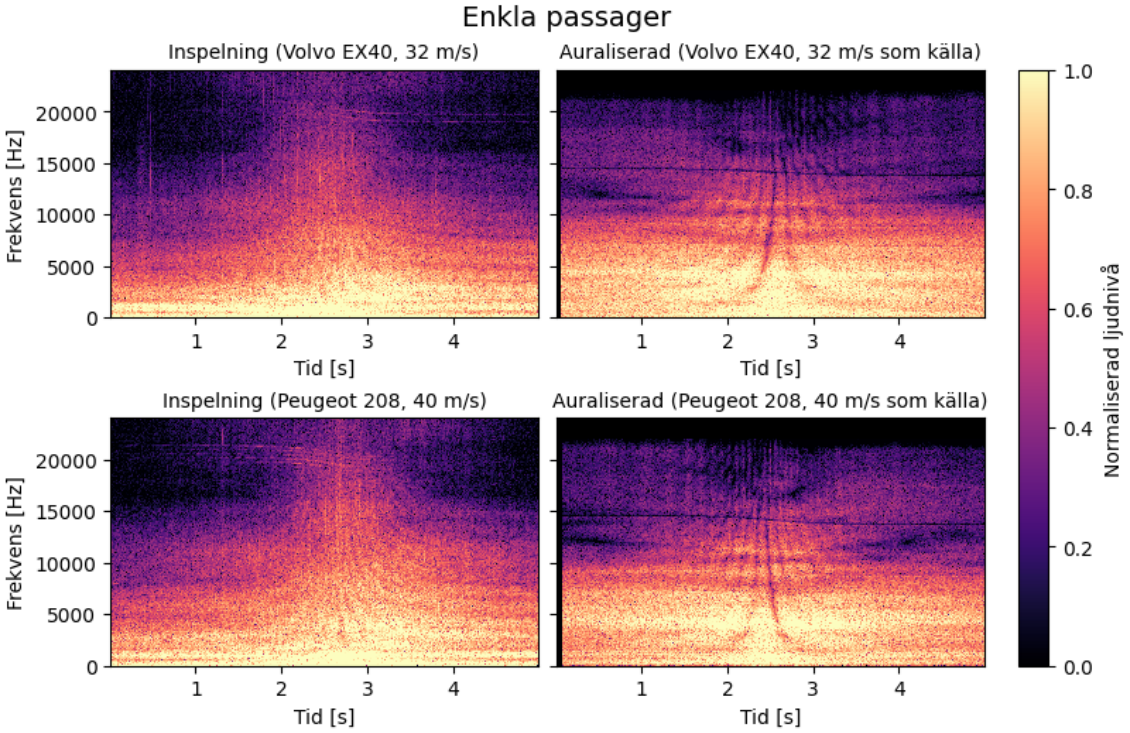
Figur 8: Renderingstid för enskilda passager beroende av samplingsfrekvens. Data kommer ifrån 60 auraliseringar av 12 olika källor. Alla auraliseringar var 4 sekunder långa och med observationsavstånd 10 m.



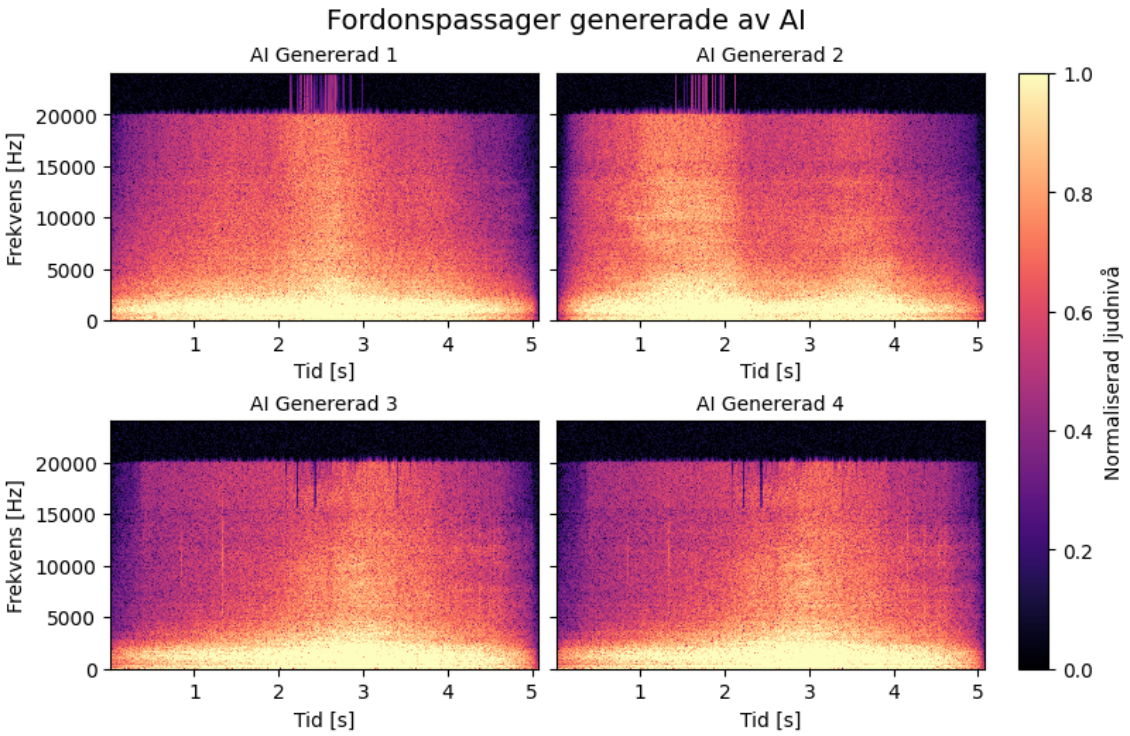
Figur 9: Renderingstid för enskilda passager beroende av auraliseringens längd i sekunder. Data kommer ifrån 72 auraliseringar av 12 olika källor. Alla auraliseringar var hade samplingsfrekvensen 44,1 kHz och med observationsavstånd 10 m.

4.2 Visualisering av ljudfiler

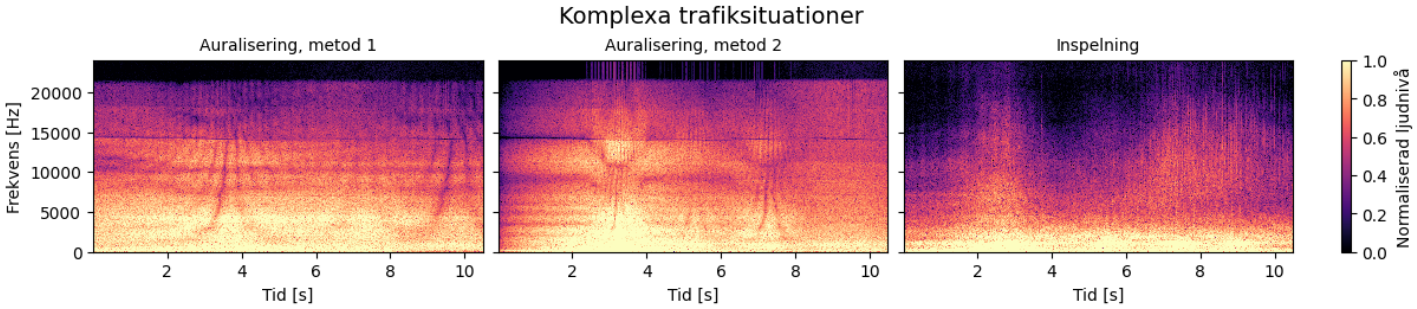
För att analytiskt kunna jämföra de auraliserade och AI-genererade ljuden med verkligheten har visualiseringar av ljudbilderna tagits fram. Ett urval av auraliserade passager och dess motsvarande ljudinspelning redovisas i figur 10. Till vänster visas passage-inspelningen som också använts som källa i auraliseringen som visas till höger. I figur 11 visas de AI-genererade fordonspassagerna skapade via ElevenLabs och i figur 12 visas spektrogrammen över de komplexa situationerna. Ljudnivån har normaliserats för att bättre visa skillnader i frekvensernas faktiska fördelning hellre än ljudnivå. Samplingsfrekvensen för inspelningar och AI-generade ljudsignaler var 48 000 Hz medan de auraliserade använde en samplingsfrekvens på 44 100 Hz.



Figur 10: Visualisering av inspelningar och auraliseringar för enskilda passager med normaliserad ljudnivå. Samplingsfrekvens för inspelning respektive auralisering var 48 000 Hz och 44 100 Hz.



Figur 11: Visualisering av AI-genererade fordonspassager med normaliserad ljudnivå. Samplingsfrekvens 48 000 Hz.



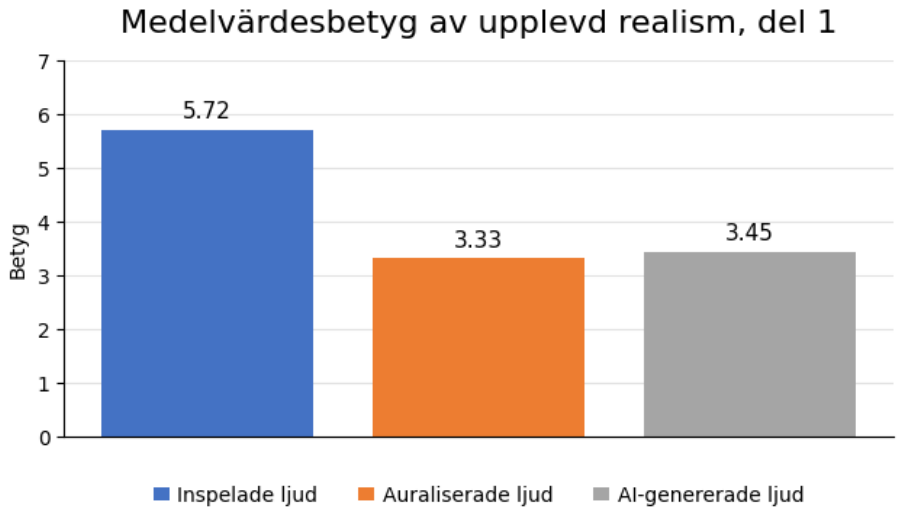
Figur 12: Visualisering av auraliserade samt inspelade komplexa situationer som användes för lyssningstestet med normaliserad ljudnivå. Samplingsfrekvensen för de auraliserade respektive inspelade var 44 100 Hz och 48 000 Hz

4.3 Analytiskt resultat av lyssningstest

I följande avsnitt presenteras de statistiska resultaten för varje del av lyssningstestet. Rådata för samtliga delar kan hittas i avsnitt A i appendix.

4.3.1 Lyssningstest - Del ett

För den första delen av testet, då enbart simpla trafiksituationer förekom, räknades medelvärdet ut per person för varje ljudkälla baserat på deras svar enligt den givna likertskalan. Här reflekterade ett betyg på 1 ett ljud som upplevts mycket orealistiskt och ett betyg på 7 ett mycket realistiskt ljud. Medelvärdet över alla deltagares betyg presenteras i figur 13.



Figur 13: Medelvärdesbetygen på samtliga ljudkällor över alla deltagare. De inspelade ljuden har ett medelvärdesbetyg på 5,71 av 7, de auraliserade 3,33 av 7 och de AI-genererade 3,45 av 7.

Ett Friedman-test utfördes på den erhållna data och resulterade i ett värde på $S = 14{,}4$. Detta jämfördes med en chitvåfördelning med två frihetsgrader och en signifikansnivå $\alpha = 0{,}05$, vilket resulterade i att $S = 14{,}4 > 5{,}991 = \chi^2_{2,\alpha=0,05}$ och att nollhypotesen om att inga systematiska skillnader mellan ljudkällorna fanns kunde förkastas. Med hjälp av detta uppskattades därefter Kendall's W till $W = 0{,}48$

Wilcoxon-testet utfördes mellan alla ljudkällor. För inspelade mot auraliserade ljud resulterade teststatistiken i $Q_1 = 3$, för inspelade mot AI-genererade ljud $Q_2 = 3$ och för auraliserade mot AI-genererade ljud $Q_3 = 58{,}5$. Dessa jämfördes alla mot det kritiska Wilcoxonvärdet $Q_{krit} = 15$.

4.3.2 Lyssningstest - Del två

Den andra delen av testet bedömde ifall deltagarna kunde skilja på inspelade och syntetiskt skapade ljud. Deltagarna hade svarsalternativen "Inspelat ljud", "Skapat ljud" och "Vet ej" på varje ljudfil. Analytiskt beräknades därefter antalet ljudkällor som korrekt identifierats, där alternativet "Vet ej" bedömdes som en inkorrekt identifiering. Resultaten presenteras i tabell 2.

Tabell 2: Andelen korrekta identifieringar av ljudklipp från inspelade, auraliserade samt AI-genererade ljud, givet i procent. Ett konfidensintervall på 95%-nivå beräknades även för varje ljudkälla.

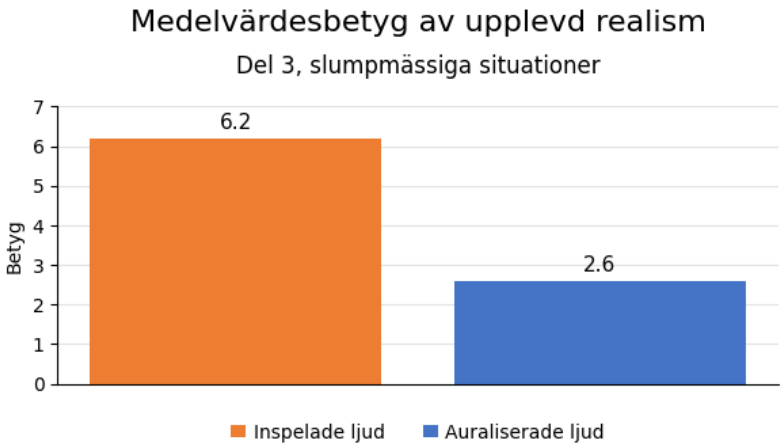
Ljudkälla	Andel korrekt identifieringar, med konfidensintervall
Inspelat	78,0±21,0%
Auraliserat	71,7±22,8%
AI-genererat	61,7±24,6%

De inspelade ljudklippen identifierades korrekt 78,0% av deltagarna, med konfidensintervall på 21,0% och ett resulterande intervall på [57,99]%. De auraliserade ljudklippen identifierades korrekt 71,7% av gångerna, med ett konfidensintervall på 22,8% och ett resulterande intervall på [48,9;94,5]%. De AI-genererade ljuden identifierades korrekt 61,7% av gångerna, med ett konfidensintervall på 26,6% och ett resulterande intervall på [37,1;86,3]%.

4.3.3 Lyssningstest - Del tre

I den sista delen av testet fick deltagarna lyssna på mer komplexa trafiksituationer och, likt den första delen, återigen betygsätta den upplevda realismen hos olika ljudklipp (på samma skala som innan). Här delades delen in i 3 underdelar, där den första innehöll slumpmässiga trafiksituationer, den andra ett specifikt fall och den tredje ett annat specifikt fall (fast med 2 olika auraliseringsmetoder). För de slumpmässiga komplexa trafiksituationerna erhölls resultaten som presenteras i figur 14.

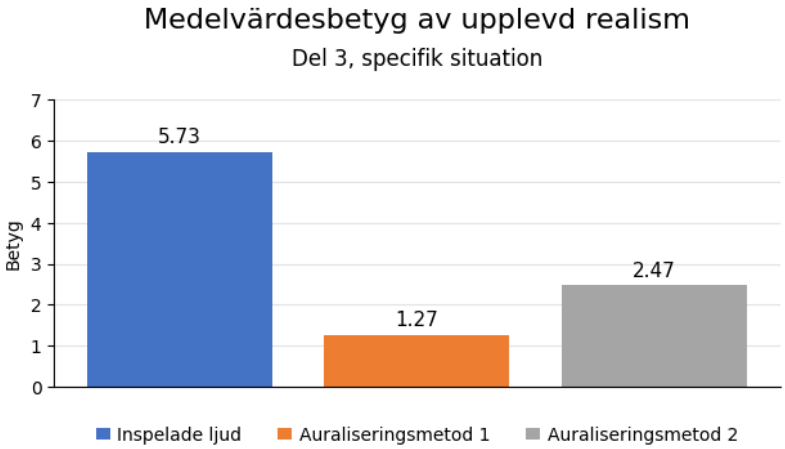
Friedman-testet på detta resultat resulterade i en teststatistik $S = 15$, jämfört med en chitvåfördelning med en frihetsgrad $\chi^2_{1,\alpha=0,05} = 3{,}841$. Kendall's W erhölls därefter med ett värde på $W = 1$ och Wilcoxon-testet resulterade i $Q = 0$, jämfört med det kritiska Wilcoxonvärdet på $Q_{krit} = 15$.



Figur 14: Medelvärdesbetygen på båda ljudkällor över alla deltagare. De inspelade ljuden har ett medelvärdesbetyg på 6,2 av 7 och de auraliserade 2,6 av 7.

Den andra underdelen, vilken behandlade en specifik trafiksituation, resulterade i exakt samma resultat som för den första underdelen. Det vill säga erhölls $S = 15$, $W = 1$ och $Q = 0$.

Den allra sista delen av testet behandlade en specifik trafiksituation, fast med ett inspelat ljud och separata ljudklipp från två olika auraliseringsmetoder. Här syftar den första auraliseringsmetoden på användandet av stereosimuleringar och den andra på användandet av binaural panorering. Resultatet för denna del presenteras i figur 15.



Figur 15: Medelvärdesbetygen på samtliga ljudkällor över alla deltagare. De inspelade ljudet har ett medelvärdesbetyg på 5,73 av 7, ljudet från den första auraliseringsmetoden 1,27 av 7 och ljudet från den andra auraliseringsmetoden 2,47 av 7.

Den resulterade teststatistiken från Friedman-testet var $S = 26{,}53$, jämfört med en chitvåfördelning med två frihetsgrader $\chi^2_{2,\alpha=0,05} = 5{,}991$. Kendall's W erhölls även med ett värde på $W = 0{,}884$.

Till sist utfördes Wilcoxon-testet mellan alla ljudkällor. För de inspelade ljuden mot den första auraliseringsmetod erhölls teststatistiken i $Q_1 = 0$, för inspelade ljud mot den andra auraliseringsmetoden $Q_2 = 0$ och för den andra mot den första auraliseringsmetoden $Q = 4{,}5$. Dessa jämfördes alla mot det kritiska Wilcoxonvärdet $Q_{krit} = 25$.

5 Diskussion

Utifrån de erhållna resultaten kunde samtliga metodikers kvalité för att syntetisera trafikbuller analyseras. I följande avsnitt bedöms processen för framställningen av de auraliserade ljudfilerna, skillnader mellan de olika ljudfilernas framtagna spektrogram och de statistiska resultaten från lyssningstestet. Därefter diskuteras möjliga förbättringsområden för metoden och avslutas med en sammanfattande slutsats om projektet i sin helhet.

5.1 Prestanda

Syftet med detta projekt var att undersöka hur auraliseringsmetoderna skulle kunna användas i en automatiserad process för att ta fram ljudbilder av en godtycklig plats i en stadsmiljö. Även om omfattningen av projektet endast täcker en utvald plats för komplex trafiksituation är det intressant att kolla på i vilken utsträckning denna metod för mer komplexa situationer skulle kunna skalas. För den trafiksituation som undersöks i rapporten krävs endast auraliseringar av ett fåtal olika avstånd, hastigheter, och dessutom enbart i en riktning, se figur 4.

En metod för att kunna sammansätta godtycklig situation är att rendera ut ett dataset av passager i förväg för olika avstånd, hastigheter, vinklar och fordonstyper. En enkel approximation av detta datasets storlek utfördes med följande antaganden. Avstånd mellan 5 och 100 m med 2 m steg, hastigheter mellan 10 och 120 km/h med 5 km/h steg, 3 fordonstyper (väldigt förenklat med lätt, medel och tung trafik), och vinklar givet i 100 steg, med varierande mellanrum följande människans hörsel som har känslighet på ca 2 grader framåt och 20 grader åt sidan. Det blir då ca 150 000 olika passager, vilket med de resultat för renderingstid och filstorlek blir totalt ca 200 timmar renderingstid och 100 GB ljudfiler. Detta resultat gör det inte helt orimligt att utifrån dagens beräkningskapacitet i datorer kunna göra en implementering på detta sätt.

5.1.1 Renderingstid av auraliseringar

För auraliseringar med samplingsfrekvens 44 100 Hz, som är standard för digitalt ljud och lämpar sig mycket bra för det hörbara spektrumet, är renderingstiden generellt något kortare än auraliseringslängden. Detta betyder inte auraliseringarna skulle kunna göras i realtid då processen inte sker i ordningen av tidssteg.

De mest trafikerade gatorna i Göteborg har i medel ca 0,4 passager per sekund och riktning. Detta skulle motsvara 8 enskilda passager i ett 10 sekunder långt ljudklipp vilket skulle ta maximalt 80 sekunder i renderingstid av källor vid maximal täckning.

5.2 Visualiserade ljudfiler

Vid analys av de visualiserade ljudfilerna kan ses att det som främst utmärker de spektrala skillnaderna mellan de AI-genererade i figur 11 och de inspelade samt auraliserade signalerna i figur 10 är det betydligt mer homogena spektrala innehållet i de AI-genererade exemplen. Detta resulterar i mindre kontrast och struktur av ljudet och transienterna framträder därmed inte lika mycket som i de inspelade eller de auraliserade ljudsignalerna. I stort sett existerar ingen energi över 20 000 Hz bortsett från ett fåtal artefakter i anslutning till transienterna, vilket kan ses på bild ”AI Genererad 1” och ”AI Genererad 2” i figur 11. Denna abrupta avskärningen av frekvenser som kan ses i alla av de AI-genererade bilderna i spektret beror sannolikt på begränsningar i renderingsprocessen, eller medveten avskärning kopplat till det mänskliga hörselomfånget som slutar runt 20 000 Hz, vid skapandet av ljudsignalen. Detta skiljer sig visuellt tydligt ifrån de inspelade exemplen där energin istället är tydligt koncentrerad kring transienterna och avtar gradvis över det resterande frekvenspektrumet utan några abrupta avskärningar.

Spektrogrammen för de auraliserade passagerna samt komplexa situationerna i figur 10 och 12 visar karaktäristiska likheter med de AI-genererade exemplen, men har dock en jämnare energifördelning samt tydligare transienter. Spektrografiskt tyder detta på ett mer distinkt uttryck än hos de AI-genererade signalerna även om resultaten fortfarande har påtagliga skillnader ifrån den verkliga inspelningen. Det kan även ses att de auraliserade ljudsignalerna har, likt de AI-genererade ljudsignalerna, en avskärningsfrekvens vid ca 22 000 Hz, som i detta fall mer tydligt kan kopplas till Nyquistfrekvensen då samplingsfrekvensen var 44 100 Hz. Detta samband till avskärningen stämmer inte överens på samma vis för de AI-genererade signalerna som hade en samplingsfrekvens på 48 000 Hz.

5.3 Lyssningstest

Utifrån de den statistiska analysen av lyssningstestet var de inspelade ljuden konsekvent rangordnade högst över alla testets delar samt hade den högsta andelen korrekta identifieringar av deltagare. I följande avsnitt tolkas de statiska resultaten från lyssningstesterna del för del.

5.3.1 Lyssningstest - Del ett

I den första delen av testet, då ljudklipp av simpla passager från varje ljudkälla var med, erhöll Friedman-testet ett $S = 14{,}4$, vilket resulterade i att $S \geq \chi^2_{2,\alpha=0,05}$. Detta innebär att systematiska skillnader i rankningar av ljud från de olika ljudkällorna fanns. Därav utfördes en beräkning av Kendall's W, vilket resulterade i $W = 0{,}48$. Detta innebär att överensstämmelsen mellan deltagarna i hur olika ljudkällor betygsattes var måttligt stark, men att variationer fortfarande finns. Det fanns

alltså tendenser hos deltagarna att rangordna de olika ljudkällorna som de gjorde, men var inte helt överens om exakt hur.

I slutändan utfördes Wilcoxon-testet mellan varje ljudkälla. Då de inspelade ljuden jämfördes mot de auraliserade samt AI-genererade ljuden erhölls samma resultat, $Q_{1,2} = 3$ vilket jämfört med det kritiska värdet $Q_{krit} = 15$ resulterade i att $Q_{1,2} \leq Q_{crit}$ och att nollhypotesen om att inga signifikanta skillnader mellan ljudkällorna fanns kunde förkastas. Här tolkas istället det låga värdet på $Q_{1,2}$ som att de inspelade ljuden systematiskt fick signifikant högre betyg än de auraliserade och AI-genererade ljuden.

Då Wilcoxon-testet utfördes mellan de auraliserade och AI-genererade ljuden erhölls istället $Q_3 = 58{,}5$, vilket jämfört med det kritiska värdet $Q_{krit} = 15$ resulterade i att $Q_3 \geq Q_{krit}$. Detta kan tolkas som att ingen signifikant skillnad i hur ljudkällorna rangordnas kunde uppskattas.

Sammanfattningsvis var alltså de inspelade ljuden av simpla trafiksituationer systematisk rangordna signifikant högre än de auraliserade samt AI-genererade ljuden. Ingen upplevd skillnad mellan de auraliserade och AI-genererade ljuden noterades.

5.3.2 Lyssningstest - Del två

I den andra delen av lyssningstestet fick deltagarna återigen lyssna på ljud från samtliga ljudkällor, men svarade enbart om de trodde att ljudet var inspelat eller skapat. Alternativet "Vet ej" fanns också som ett svarsalternativ, men bedömdes i analysen som en felaktig identifiering av ett ljud, då deltagaren helt enkelt inte kunde bedöma om ljudet var verkligt eller inte.

De inspelade ljuden identifierades korrekt med störst framgång 78% av gångerna över samtliga deltagare. Konfidensintervallet på en 95%-nivå uppskattades till ±21,0%, vilket resulterar i ett intervall på [57,99]%. Då intervallet ligger över 50% kan identifiering av inspelade ljud tolkas som att vara bättre än slumpen och relativt säkert kunna identifieras av en lyssnare.

De auraliserade ljuden identifierades korrekt 71,7% av gångerna, med ett konfidensintervall på $\pm 22,8\%$ och det resulterande intervallet [48,9;94,5]%. Då konfidensintervallet innehåller värden under 50% går det inte att utesluta slumpfaktorn i identifieringen, men med ett relativt stort konfidensintervall är detta svårt att avgöra.

De AI-genererade ljuden identifierades korrekt 61,7%, med ett konfidensintervall på ±24,6% och ett resulterande intervall på [37,1;86,3]%. Återigen innehåller konfidensintervallet värden under 50%, vilket resulterar i att slumpfaktorn inte kan uteslutas i identifieringen av ljuden.

Sammanfattningsvis kunde ändå deltagarna med hög säkerhet identifiera ifall ett ljud var skapat eller inte. Däremot är konfidensintervallen på samtliga ljudkällor relativt stora, vilket indikerar på en viss variation hos deltagarnas förmåga under testet. Den största slutsatsen från denna del är att inspelade ljud kan identifieras med stor säkerhet och att de auraliserade samt AI-genererade ljuden kan identifieras med relativt stor framgång, men att slumpfaktorn ändå kan ha en avgörande verkan.

5.3.3 Lyssningstest - Del tre

I den slutgiltiga delen av testet med komplexa trafiksituationer, bestående av tre delar, fick deltagarna först lyssna på slumpmässiga trafiksituationer och sedan två specifik fall. Den största skillnaden mellan detta och resterande delar av testet var att AI-genererade ljud inte var med, på grund av att ingen tillgänglig AI var kapabel till att skapa trafiksituationer med fler än ett fordon.

I den första delen bestod ljuden av slumpmässiga komplexa trafiksituationer med bidrag från både inspelade och auraliserade ljudklipp. Den resulterande teststatistiken från Friedman-testet blev $S = 15$, vilket resulterade i att $S \geq \chi^2_{1,\alpha=0,05}$. Därav fanns systematiska skillnader mellan de inspelade och auraliserade ljuden, vilket stärktes med det erhållna Kendall's W på $W = 1$ som innebär en total överenskommelse mellan deltagarna. Med ett erhållet $Q_1 = 0$ från Wilcoxon-testet, jämfört med det kritiska värdet $Q_{krit} = 15$, resulterade detta test i att de inspelade ljuden systematiskt fick signifikant högre betyg än de auraliserade med total överenskommelse mellan deltagarna.

I den andra delen lyssnade deltagarna på ett inspelat och auraliserat ljud, där det auraliserade ljudet försökts återskapa det inspelade ljudets trafiksituation i så hög grad som möjligt. Resultatet var däremot det samma som i delen innan, med $S = 15$, $W = 1$ och $Q_2 = 0$, vilket tolkas exakt likadant. Det inspelade ljudet blev systematiskt högre betygsatt med total överenskommelse mellan deltagarna.

Den sista delen innehöll en inspelad trafiksituation och två olika typer av auraliserade ljud, vilka båda försökts återskapa det inspelade ljudet. Ett Friedman-test resulterade i $S = 26{,}53$, vilket jämfördes med $\chi^2_{2,\alpha=0,05} = 5{,}991$. Då $S > \chi^2_{2,\alpha=0,05}$ finns systematiska skillnader mellan betygssättningen av ljud och det resulterande Kendall's W på $W = 0{,}884$ berättar att deltagarna var mycket överens med sina betygsättningar.

Efter att Wilcoxon-testet utförts mellan de inspelade ljuden samt båda auraliserade ljuden erhölls teststatistiken $Q = 0$, jämfört med det kritiska värdet $Q_{krit} = 25$ vilket innebar att de inspelade ljuden signifikant betygssatts bättre än de auraliserade. Däremot gav testet ett $Q = 4,5$ när de två auraliserade ljuden ställdes mot varandra, med resultatet av att den andra auraliseringsmetoden som betygssatts signifikant mycket bättre än den första.

Sammanfattningsvis uppfattades de inspelade ljuden i mycket högre grad som realistiska än de auraliserade. Däremot uppfattades den andra auraliseringsmetoden som signifikant mer realistisk än den första, vilket ger oss bekräftelsen att den andra auraliseringsmetoden är betydligt bättre än den första. Däremot var den binaurala panoreringen som användes i den andra auraliseringsmetoden ett manuellt och perceptuellt inlägg, vilket även motiverar att en helt automatiserad process ännu inte ger det bästa resultatet.

5.3.4 Lyssningstester - Kommentarer från deltagare

Under lyssningstestets alla delar fanns möjligheten för deltagarna att lämna kommentarer om varför man betygssatt eller identifierat ett visst ljudklipp som man gjort. Dessvärre kunde inte specifika

kommentarer kopplas till specifika ord på grund av den slumpmässiga ordning varje ljudklipp kom i för varje deltagare. Data över i vilken ordning en specifik deltagare fått ljudklippen i kunde inte hittas och därav kan kommentarerna enbart tolkas övergripande. Dessutom bestod de flesta av kommentarerna endast av ett kort beskrivande ord, vilket i vissa fall gjorde det svårt att avgöra exakt vad deltagaren menade.

De mest frekventa kommentarerna om varför man rangordnat ett ljud lägre eller identifierat det som syntetiskt skapat var att det uppfattades som blött, brusigt eller ha ett orealistiskt ljudbidrag från fordonets däck. Kommentarer som syftar på att ett ljud lät blött eller brusigt skulle kunna tyda på att vissa frekvenskomponenter i ljudklippen avviker med för stor grad från vad som förväntas av ett realistiskt ljud. I det fall det var de auraliserade ljuden som upplevts blöta eller brusiga skulle detta antagligen kunna vara en följd av att auraliseringsprocessen av inspelade ljud inte varit optimal. Om detta beror på att själva inspelningar inte utfördes under tillräckligt bra förhållanden eller att auraliseringen måste optimeras är dock oklart.

Att ljudklipp skulle innehålla orealistiska ljudbidrag från fordonens däck skulle kunna bero på flera faktorer. En avgörande anledning skulle kunna vara ifall de inspelade fordonen hade dubbdäck eller inte. Dubbdäck bidrar med en betydligt skild ljudbild jämfört med fordon utan dessa. Då inspelningar utfördes både under en period då dubbdäck och senare sommardäck användes kan den insamlade data avsevärt skilja sig åt beroende på just valet av typ av däck på inspelade fordon. Denna faktor beräknades inte in under auraliseringar och framtida modeller vilka hanterar dubbdäck skulle möjligtvis behöva undersökas och tillämpas.

5.4 Förbättringsområden

Vid diskussion och reflektion över resultatet av det genomförda projektet framfördes i efterhand en del möjliga förbättringsområden. Tillvägagångssättet av metoden var till stor del överbelastad av datainsamling och mättillfällen som var tidskrävande samt förhindrade möjligheterna att skapa auraliseringar i ett tidigare skede av projektet. Denna datainsamling ansågs även i efterhand vara till överdriven mängd, otillräcklig kvalitet, samt att inte vara så väsentlig för projektets syfte och mål än vad inledningsvis antagits. Den stora mängden data gjorde även att organiseringen och genomgången av erhållen data krävde en lång tidsåtgång. I slutändan konstateras därav att auraliseringsförsöken via koden borde bearbetats och genomförts tidigare samt även att sammanställningen i *Logic Pro X* borde genomförts tidigare då det hade kunnat hjälpa att eventuellt förbättra koden för auralisering och att implementera flera ljudkällor samtidigt. Genom att tidigare kolla på vad som gjordes i sammanställningen via *Logic Pro X* hade projektet kunnat lägga större fokus på att efterlikna de stegen som genomfördes i programmet och sedan försöka efterlikna detta i koden för auralisering.

En faktor som med stor sannolikhet påverkat kvaliteten på källorna och i sin tur resultatet negativt var ett relativt fuktigt vägunderlag för fordonen under en stor del av datainsamlingen. För att undvika detta hade inspelningarna med stor fördel utförts under tider med mer gynnande väderförhållanden. Detta hade då kunnat bidra till en mer konsekvent datainsamling samt ett bättre slutgiltigt resultat.

Slutligen finns även utrymme för att helt lyckas auralisera spårvagnar. På grund av delvis dess längd fungerade modellering av spårvagnar som punktkällor på samma sätt som övriga fordon mycket dåligt. Spårvagnar har dessutom en betydligt mer komplicerad och inkonsistent ljudbild med hjul som slår mot rälsen, acceleration från elmotorn i flera steg och gnissel i vagnarna. För att kunna få en någorlunda realistisk auralisering av spårvagnar behöver antagligen denna modelleras som flera källor utspridda över spårvagnen.

5.5 Slutsats

Sammanfattningsvis upplevdes varken de auraliserade eller AI-genererade ljudklippen realistiska jämfört med de faktiska inspelningarna. I det genomförda lyssningstestet rangordnas de inspelade ljudklippen konsekvent högst som mest realistiska och deltagarna kunde med relativt stor säkerhet avgöra ifall ett ljudklipp var en inspelning eller syntetiskt skapat. Detta är även ett rimligt utfall då de erhållna spektrogrammen signifikant skilde sig åt mellan inspelningar samt de syntetiskt skapade ljuden. Därav finns fortfarande ett stort utrymme för utvecklingen att ta fram syntetiskt skapade ljud.

I syfte att utveckla dessa metodiker vidare och uppnå mer tillfredsställande resultat hade delar av projektet som exempelvis inspelningstillfällen samt auraliseringsmetodik behövt ses över. Detta stärks ytterligare av att den auraliseringsmetod med inslag av manuell och perceptuell redigering upplevdes i signifikant högre grad mer realistisk än auraliseringsmetoden som inte gjorde det. Därav bör vidare utveckling av modellen och inslag av exempelvis binaural panorering ses över för att ytterligare stärka metoden och de slutliga resultaten.

Däremot uppskattas ett större projekt och möjligheter för att skapa ett större bibliotek av auraliserade ljudfiler i syfte att bidra till framtida mer automatiserade ljudsynteser av trafikmiljöer vara ytterst möjligt. Den tid och det lagringsutrymme som skulle krävas för att erhålla ett sådant bibliotek tycks vara mycket rimligt i förhållande till vad det skulle kunna bidra till i framtida projekt.

Dessutom finns flera möjligheter att fokusera mer på auraliseringen av flera typer av fordon, som exempelvis spårvagnar. Detta lyckades inte genomföras under projektets gång, men anses vara högst troligt att vara genomförbart i mån av mer tid och arbete.

Referenser

[1] World Health Organization m. fl. "Environmental noise guidelines for the European region". I: (2018).

[2] European Environment Agency. *Environmental Noise in Europe — 2025*. EEA Report TH-01-25-026-EN-N. Copenhagen: European Environment Agency, 2025. DOI: 10.2800/1181642. Tillgänglig från: https://www.eea.europa.eu/en/analysis/publications/environmental-noise-in-europe-2025.

[3] Regeringskansliet, Klimat- och näringslivsdepartementet. *Förordning (2004:675) om omgivningsbuller*. 2004. Tillgänglig från: https://rkrattsbaser.gov.se/sfst?bet=2004:675.

[4] Naturvårdsverket. *Undersökningar av antalet exponerade för buller*. Granskad: 2025-07-14. Tillgänglig från: https://www.naturvardsverket.se/amnesomraden/buller/undersokningar-av-antalet-exponerade-for-buller/ (Hämtad: 2026-02-09).

[5] Mats E. Nilsson, Mikael Andéhn och Paulina Leśna. "Evaluating roadside noise barriers using an annoyance-reduction criterion". I: *The Journal of the Acoustical Society of America* 124.6 (dec. 2008), s. 3561–3567. DOI: 10.1121/1.2997433.

[6] Michael Vorländer. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer Berlin Heidelberg, 2008. ISBN: 978-3-540-48830-9. DOI: 10.1007/978-3-540-48830-9.

[7] Mats E. Nilsson m. fl. *LISTEN Auralization of Urban Soundscapes*. Final report to the Knowledge Foundation. 30 sept. 2011. Tillgänglig från: https://publications.lib.chalmers.se/records/fulltext/232736/local_232736.pdf (Hämtad: 2026-02-06).

[8] *Road Vehicles: Determination of Noise Emission*. Nordtest Method NT ACOU 116. Oslo, Norway: Nordtest, nov. 2004.

[9] International Organization for Standardization. *Acoustics - Attenuation of sound during propagation outdoors - Part 1: Calculation of the absorption of sound by the atmosphere*. Tillgänglig från: https://www.iso.org/obp/ui/en/#iso:std:iso:9613:-1:ed-1:v1:en (Hämtad: 2026-02-13).

[10] Jonny Nordling Carl och Österman. *Physics Handbook for Science and Engineering*. 9th. Studentlitteratur, 2020.

[11] R. Nota, R. Barelds och D. van Maercke. *Engineering method for road traffic and railway noise after validation and fine-tuning*. Technical Report HAR32TR-040922-DGMR20. EC under the Information Society och Technology (IST) Programme, 2005.

[12] William Moebs, Samuel J. Ling och Jeff Sanny. *University Physics Volume 1*. Section: https://openstax.org/books/university-physics-volume-1/pages/17-7-the-doppler-effect. Houston, Texas: OpenStax, 2016. Tillgänglig från: https://openstax.org/books/university-physics-volume-1/pages/1-introduction (Hämtad: 2026-02-08).

[13] Maschen. *Relativistic Doppler Effect [Image]*. https://commons.wikimedia.org/wiki/File:Relativistic_Doppler_Effect.svg. Creative Commons CC0 1.0 Universal Public Domain Dedication. 2015. Tillgänglig från: https://commons.wikimedia.org/wiki/File:Relativistic_Doppler_Effect.svg (Hämtad: 2026-02-08).

[14] National Instruments. *Microphone Handbook: Types, Components & Testing*. Tillgänglig från: https://www.ni.com/en/shop/data-acquisition/sensor-fundamentals/measuring-sound-with-microphones/microphone-handbook.html (Hämtad: 2026-05-05).

[15] Stanford University. *WAVE PCM soundfile format*. Tillgänglig från: http://soundfile.sapp.org/doc/WaveFormat/ (Hämtad: 2026-05-12).

[16] The University of Queensland. *Nyquist Conditions*. Tillgänglig från: https://imb.uq.edu.au/research/facilities/microscopy/training-manuals/microscopy-online-resources/image-capture/nyquist-conditions (Hämtad: 2026-05-13).

[17] National Instruments. *Understanding FFTs and Windowing*. Tillgänglig från: https://download.ni.com/evaluation/pxi/Understanding%20FFTs%20and%20Windowing.pdf (Hämtad: 2026-05-12).

[18] Myles Hollander, Douglas A. Wolfe och Eric Chicken. *Nonparametric Statistical Methods*. 3. utg. John Wiley & Sons, 2013. ISBN: 9780470387375.

[19] Lennart Råde och Bertil Westergren. *Mathematics Handbook for Science and Engineering*. 6th. Studentlitteratur, 2019.

[20] Pierre Legendre. "Coefficient of Concordance". I: *Encyclopedia of Research Design*. Utg. av Neil J. Salkind. Los Angeles: SAGE Publications, 2010, s. 164–169. ISBN: 9781412961271.

[21] Nicolas Riche m. fl. "A Study of Parameters Affecting Visual Saliency Assessment". I: *arXiv preprint arXiv:1307.5691* (2013). Tillgänglig från: https://arxiv.org/abs/1307.5691.

[22] University of Saskatchewan Distance Education Unit. *Wilcoxon Signed-Rank Test Critical Values Table*. 2020. Tillgänglig från: https://www.saskoer.ca/app/uploads/sites/313/2020/11/Wilcoxon-Signed-Rank-Test-Critical-Values-Table.pdf (Hämtad: 2026-05-05).

[23] Avijit Hazra. "Using the confidence interval confidently". I: *Journal of Thoracic Disease* 9.10 (2017). ISSN: 2077-6624. Tillgänglig från: https://jtd.amegroups.org/article/view/16406.

[24] L. A. Orawo. "Confidence Intervals for the Binomial Proportion: A Comparison of Four Methods". I: *Open Journal of Statistics* 11.5 (2021), s. 806–816. DOI: 10.4236/ojs.2021.115047. Tillgänglig från: https://doi.org/10.4236/ojs.2021.115047.

[25] OptimizerAI. *About OptimizerAI*. https://www.optimizerai.xyz/about. [Online; hämtad 9 apr. 2026].

[26] ElevenLabs. *About*. 2026. Tillgänglig från: https://elevenlabs.io/about (Hämtad: 2026-05-13).

[27] ElevenLabs. *ElevenLabs - Sound Effects*. https://elevenlabs.io/sv/sound-effects. [Online; hämtad 9 apr. 2026].

A Data från lyssningstester

I följande avsnitt presenteras all data från de delar av lyssningstestet där ljud betygsattes på en likertskala mellan ett och sju baserat på upplevd grad av realism. Här korresponderar ett betyg på 1 ett mycket orealistiskt ljud och ett betyg på 7 ett extremt realistiskt ljud. Data har sammanställts utefter medelbetyget en deltagare gav till ljuden från en specifik ljudkälla.

I tabell 3 presenteras data från den första delen av lyssningstestet, där enbart enkla passager var med.

Tabell 3: Erhållna resultat utifrån betygssättning av upplevd realism av ljudklipp av simpla trafiksituationer från inspelade, auraliserade samt AI-genererade ljud. Alla värden representerar det medelvärdesbetyg för varje ljudkälla som varje deltagare gav.

Deltagare	Betyg på inspelade ljud	Betyg på auraliserade ljud	Betyg på AI-genererade ljud
1	4,75	3,25	3
2	4,5	5,25	3
3	4,5	1,5	4
4	7	3,25	5
5	6,5	3,75	3,25
6	5,25	3,75	2,5
7	6,75	6,5	5,25
8	6,25	2,5	6,5
9	6	4,75	1,5
10	6,75	3	4,25
11	6	1,5	2
12	6	1,5	3,5
13	7	4	3,5
14	3,25	3	3,5
15	5,25	2,5	1,25

I tabell 4 presenteras data från den sista delen av lyssningstestet, då komplexa men slumpmässiga trafiksituationer var behandlades.

Tabell 4: Erhållna resultat utifrån betygssättning av upplevd realism av ljudklipp av komplexa och slumpmässiga trafiksituationer från inspelade samt auraliserade ljud. Alla värden representerar det medelvärdesbetyg för varje ljudkälla som varje deltagare gav.

Deltagare	Betyg på inspelade ljud	Betyg på auraliserade ljud
1	6,33	1,67
2	6,33	3
3	5,67	3,33
4	7	1,67
5	6,33	3
6	4,33	3,33
7	6	2,33
8	6,67	2
9	6,67	2,33
10	6,33	2,67
11	6,67	4
12	5,67	2
13	6,67	3
14	6,67	2
15	6	2,67

I tabell 5 presenteras data från den första specifika komplexa trafiksituationen, vilken bestod av ett inspelat och ett auraliserat ljud.

Tabell 5: Erhållna resultat utifrån betygssättning av upplevd realism av den första specifika komplexa trafiksituationen baserad på en inspelning samt ett auraliserat ljud. Alla värden representerar det medelvärdesbetyg för varje ljudkälla som varje deltagare gav.

Deltagare	Betyg på inspelat ljud	Betyg på auraliserat ljud
1	4	1
2	6	4
3	6	3
4	5	2
5	6	4
6	5	3
7	7	4
8	6	1
9	7	1
10	7	6
11	7	6
12	7	1
13	7	2
14	4	2
15	3	1

I tabell 6 presenteras data från den allra sista delen av lyssningstestet, då den andra specifik trafiksituation baserad på ett inspelat klipp följt av auraliserade ljud från två olika auraliseringsmetoder var med.

Tabell 6: Erhållna resultat utifrån betygssättning av upplevd realism av ljudklipp av en specifik komplex trafiksituation från ett inspelat ljud samt två olika auraliseringsmetoder. Alla värden representerar det medelvärdesbetyg för varje ljudkälla som varje deltagare gav.

Deltagare	Betyg på inspelat ljud	Betyg på auraliseringsmetod 1	Betyg på auraliseringsmetod 2
1	3	1	1
2	5	1	2
3	6	1	2
4	7	1	1
5	6	3	2
6	5	1	3
7	7	1	3
8	7	1	2
9	4	1	2
10	7	3	4
11	6	1	4
12	7	1	2
13	7	1	4
14	5	1	2
15	4	1	3

B Auralisering, Implementering i Python

Källkoden för auralisering av enskilda passager finns att tillgå via git på länken nedan: https://git.chalmers.se/andtornk/acex11-vt26-61a

C Metadata för använda käll-ljud

De källor som användes i auraliseringarna utgick från 12 inspelningar av olika fordon i olika hastigheter se tabell 7.

Hastighet	**Fordonstyp**	**Tillverkare**	**Modell**	**Produktionsår**
32,1	Personbil	VOLVO	EX40	2025
40,0	Personbil	PEUGEOT	208	2022
31,0	Personbil	SKODA	ENYAQ 60	2021
36,0	Skåpbil	OPEL	COMBO	2023
22,5	Lastbil	SCANIA	P280DB4X2MNB	2017
22,5	Personbil	TOYOTA	COROLLA	2025
19,0	Personbil	TOYOTA	COROLLA	2025
20,0	Personbil	CUPRA	BORN	2025
40,0	Skåpbil	OPEL	VIVARO	2019
37,7	Personbil	TESLA	MODEL Y	2024
31,3	Skåpbil	FORD	TRANSIT	2024
21,2	Personbil	MERCEDES-BENZ	E 300 DE	2020

Tabell 7: Fordonsdata för de pasager som användes som källor i auraliseringar.