

# Användning av mikrofonuppsättningar och beamforming i talkonsoler

IRINA ISELIND

PATRIK MÅNSSON

BACHELOR'S THESIS

DEPARTMENT OF ELECTRICAL AND INFORMATION TECHNOLOGY |

FACULTY OF ENGINEERING | LTH | LUND UNIVERSITY





# Användning av mikrofonuppsättningar och beamforming i talkonsoler

Av

Irina Iselind och Patrik Månsson

Department of Electrical and Information Technology  
Faculty of Engineering, LTH, Lund University  
SE-221 00 Lund, Sweden

## Sammanfattning

Syftet med examensarbetet var att undersöka hur en uppsättning av mikrofoner, inbyggda i en kommunikationskonsol, kan användas på avstånd för att ge användaren fördelar som close talk-lösningar erbjuder, samt bygga en prototyp av produkten. För att ta fram varianter lämpliga att provas utfördes en förstudie i form av en litteraturstudie. Denna resulterade i ett antal alternativ som skiljer sig i antal element samt placeringen. Dessa alternativ testades i examensarbetet, där ljudet spelades in via Audacity och därefter bearbetades i MATLAB. Ljudåtergivningen i de processade ljudfilerna bedömdes ifråga om taltydlighet. Utvärderingen utfördes med hjälp av MATLAB-funktioner och ett subjektiv bedömningstest. Resultatet blev att tre eller fyra mikrofoner av en specifik typ, placerade i en linjär mikrofonuppsättning på fronten av konsolen, ger bäst resultat.

## Nyckelord

Beamforming, mikrofonuppsättning, PESQ, STOI, MOS

## Abstract

The first purpose of the thesis was to investigate how a microphone array in a communication console can be used for giving the user the advantages of the close talk solutions, like a gooseneck microphone. The second purpose was to build a prototype of the product. In order to find variants suitable for testing, a preliminary investigation was carried out in form of a literature study. This resulted in a number of alternatives, that differ in number of elements and location of the microphones at the console. These options were tested, and the sound was recorded via Audacity and then processed in MATLAB. The speech reproduction in the processed audio files was analyzed in terms of speech intelligibility. The evaluation was performed using MATLAB functions and a subjective assessment test. The result was that three or four microphones of a certain type, placed in a linear microphone array on the front of the console, give the best results.

## Keywords

Beamforming, microphone array, PESQ, STOI, MOS

## Förord

Vi vill rikta ett varmt tack till alla som har hjälpt oss, genom att dela med sig av sin kunskap och komma med råd under arbetets gång. Först vill vi tacka Axis Communications AB för att vi har fått möjligheten att utvecklas på företaget. Vi vill rikta ett särskilt tack till Carl Hansson som har väglett oss genom processen med kunnighet och varmt hjärta. Vi vill även uppmärksamma vår handledare Mats Lilja för intressanta samtal, tålamod och goda råd samt examinatorn Christian Nyberg. Även bibliotekarien Berit Funke vill vi rikta ett tack till, för hennes tillmötesgående och tid. Sist men inte minst vill vi säga tack till våra familjer som har alltid funnits där för oss med stöd och uppmuntran.

Irina Iselind och Patrik Månsson

|                                                                                                 |    |
|-------------------------------------------------------------------------------------------------|----|
| Sammanfattning .....                                                                            | 1  |
| Nyckelord .....                                                                                 | 1  |
| Abstract .....                                                                                  | 2  |
| Keywords .....                                                                                  | 2  |
| Förord .....                                                                                    | 3  |
| 1 Inledning .....                                                                               | 8  |
| 1.1 Bakgrund .....                                                                              | 8  |
| 1.2 Syfte .....                                                                                 | 8  |
| 1.3 Målformulering .....                                                                        | 9  |
| 1.4 Problemformulering .....                                                                    | 9  |
| 1.5 Motivering av examensarbetet .....                                                          | 9  |
| 1.6 Avgränsningar .....                                                                         | 9  |
| 2 Teknisk bakgrund .....                                                                        | 10 |
| 2.1 Mikrofoner .....                                                                            | 10 |
| 2.1.1 Uppbyggnad .....                                                                          | 10 |
| 2.1.2 Riktningsskarakteristik .....                                                             | 13 |
| 2.1.3 Slutsats utifrån litteraturstudien .....                                                  | 15 |
| 2.2 Mikrofonuppsättning .....                                                                   | 15 |
| 2.2.1 Dimensionell placering av mikrofoner .....                                                | 15 |
| 2.2.2 Mikrofonuppsättningen utifrån konsolens geometri .....                                    | 16 |
| 2.2.3 Avstånd mellan enskilda mikrofoner .....                                                  | 16 |
| 2.2.4 Antal element .....                                                                       | 17 |
| 2.2.5 Placering .....                                                                           | 17 |
| 2.2.6 Sammanfattning av de mikrofonuppsättningsalternativ som undersöktes i examensarbetet .... | 18 |
| 2.3 Beamforming .....                                                                           | 18 |
| 2.3.1 Konventionella beamformers .....                                                          | 18 |
| 2.3.2 Adaptiva beamformers .....                                                                | 19 |
| 2.4 Verktyg för in- och uppspelning av ljud .....                                               | 19 |
| 2.4.1 Ljudkort för inspelning från mikrofonuppsättningen .....                                  | 19 |
| 2.4.2 Mixerbord för uppspelning av tal .....                                                    | 20 |
| 2.4.3 Mixerbord för uppspelning av vitt brus .....                                              | 20 |
| 2.4.4 Decibelmätare .....                                                                       | 20 |
| 2.4.5 Högtalare .....                                                                           | 21 |

|                                                                               |    |
|-------------------------------------------------------------------------------|----|
| 2.4.6 Hörlurar .....                                                          | 21 |
| 2.4.7 Mikrofonerna i examensarbetet .....                                     | 21 |
| 2.5 Digitala bearbningsverktyg .....                                          | 22 |
| 2.5.1 MATLAB .....                                                            | 22 |
| 2.5.2 Audacity .....                                                          | 22 |
| 2.6 Utvärderingsverktyg .....                                                 | 23 |
| 2.6.1 STOI .....                                                              | 23 |
| 2.6.2 PESQ .....                                                              | 23 |
| 3 Metod .....                                                                 | 24 |
| 3.1 Arbetets faser .....                                                      | 24 |
| 3.1.1 Litteraturstudien. ....                                                 | 24 |
| 3.1.2 Prototypframställning.....                                              | 24 |
| 3.1.3 Testning.....                                                           | 25 |
| 3.1.4 Utvärdering.....                                                        | 25 |
| 3.1.5 Sammanställning .....                                                   | 25 |
| 3.1.6 Rapportskrivning .....                                                  | 25 |
| 3.2 Kommunikation .....                                                       | 25 |
| 3.3 Val av metoder .....                                                      | 26 |
| 3.4 Val av verktyg .....                                                      | 26 |
| 3.5 Sätt att inhämta information .....                                        | 26 |
| 3.6 Testuppställning och ljudinspelning .....                                 | 27 |
| 3.6.1 Ljudnivåer .....                                                        | 27 |
| 3.6.2 Val av ljud .....                                                       | 27 |
| 3.6.3 Testuppställning .....                                                  | 28 |
| 3.6.4 Informationsbearbetning.....                                            | 36 |
| 3.6.5 Subjektivt lyssnartest .....                                            | 40 |
| 3.6.6 Ytterligare testinspelning med bäst lämpade mikrofonuppsättningar ..... | 40 |
| 3.7 Källkritik.....                                                           | 41 |
| 4 Analys.....                                                                 | 42 |
| 4.1 Mikrofonval .....                                                         | 42 |
| 4.2 Mikrofonuppsättningen.....                                                | 42 |
| 4.3 Resultat från databearbetning .....                                       | 42 |
| 4.4 Kravspecifikation.....                                                    | 46 |

|                                                                       |    |
|-----------------------------------------------------------------------|----|
| 4.5 Kostnad.....                                                      | 46 |
| 4.6 Valens betydelse för resultatet.....                              | 46 |
| 4.6.1 MATLAB vs Python.....                                           | 46 |
| 4.6.2 Brusreducering.....                                             | 47 |
| 4.7 Problem och lösningar.....                                        | 47 |
| 4.7.1 Utvärderingskort.....                                           | 47 |
| 4.7.2 Avsaknad av manual till rätt version av utvärderingskort.....   | 47 |
| 4.7.3 Oklarheter kring val av programvara.....                        | 47 |
| 4.7.4 Utvärderingskort kan bara ta in signal från två mikrofoner..... | 47 |
| 4.7.5 Avsaknad av komponenter.....                                    | 48 |
| 4.7.6 Elektriskt brus.....                                            | 48 |
| 5 Resultat.....                                                       | 49 |
| 5.1 Antal mikrofoner.....                                             | 49 |
| 5.2 Placering.....                                                    | 49 |
| 5.3 MOS-test.....                                                     | 50 |
| 5.4 Kostnader.....                                                    | 50 |
| 6 Slutsats.....                                                       | 51 |
| 6.1 Reflektion över etiska aspekter.....                              | 51 |
| 6.1.1 Försiktighet vid fotografering.....                             | 51 |
| 6.1.2 Integritet.....                                                 | 51 |
| 6.1.3 Miljö.....                                                      | 52 |
| 6.2 Framtida utvecklingsmöjligheter.....                              | 52 |
| 6.2.1 Mikrofonuppsättningens geometri.....                            | 52 |
| 6.2.2 Avskärmning.....                                                | 52 |
| 6.2.3 Algoritmer för beamforming.....                                 | 53 |
| 6.2.4 Ytterligare efterbehandling av ljud.....                        | 53 |
| 6.2.5 Fler inspelningar.....                                          | 56 |
| 6.2.6 Användning av MEMS-mikrofoner.....                              | 57 |
| 6.2.7 Utvärderingskort.....                                           | 57 |
| 6.2.8 Sammanfattning av framtida utvecklingsmöjligheter.....          | 58 |
| 7 Terminologi.....                                                    | 59 |
| 8 Källförteckning.....                                                | 60 |
| Appendix A: Polärdiagram.....                                         | 65 |

|                                                                                                |    |
|------------------------------------------------------------------------------------------------|----|
| Appendix B: MATLAB-kod för bearbetning av ljudfiler.....                                       | 69 |
| Appendix C: Blankett för poängsättning vid MOS-test .....                                      | 81 |
| Appendix D: Innebörden av ljudfilernas numrering vid MOS-test.....                             | 83 |
| Appendix E: Individuella svar i MOS-testet.....                                                | 84 |
| Appendix F: Grafisk visualisering av filter för förarbetning av ljudfil till munsimulator..... | 86 |

# 1 Inledning

## 1.1 Bakgrund

Examensarbetet genomfördes i samarbete med Axis Communications AB. Företaget är världsledande på marknaden för nätverkskameror och video-encoders. Utöver dessa produkter utvecklar Axis Communications AB även passagekontroll, nätverksanslutna högtalare och IP-videoporttelefoner. Företaget har sitt huvudkontor i Lund och har över 3000 anställda i världen.

Användningen av mikrofoner, där talaren finns nära mikrofonen, är vanligt av flera skäl, till exempel välfungerande överföring av tal och användarvänlighet. Ett headset eller en svanhalsmikrofon är dock skrymmande, känslig för slitage och ibland opraktisk. Examensarbetet löser problemet med att befintliga close talk-lösningar kan upplevs som opraktiska och slitagekänsliga. En sådan lösning erbjuder samma fördelar som close talk-mikrofoner och samtidigt möjliggör ett längre avstånd mellan ljudkällan och ljuddetektorn.

En sökning efter liknande examensarbete genomfördes i databasen UppSök. Sökningen på termen "microphone array" gav 36 träffar, varav 6 valdes ut som intressanta. Sökningen på termen "beamforming" gav 57 träffar, varav ett examensarbete var intressant, dock har det arbetet funnits redan vid sökningen på "microphone array". De genomförda sökningarna visade att det inte finns något examensarbete som avhandlar exakt samma ämne. Arbeten som var intressanta hanterade microphone array i konferenssammanhang, dock är det ett ämne som bara tangerar området för detta examensarbete. Uppsatserna som var intressanta studerades utifrån referenssammanhang.

## 1.2 Syfte

Syftet med arbetet är att undersöka hur en uppsättning av mikrofoner kan användas på avstånd för att, genom beamforming, ge användaren fördelar som en close talk-mikrofon erbjuder, och bygga en prototyp av produkten. Axis Communications AB är intresserat av att undersöka tekniken för att kunna vidareutveckla sitt kommunikationssystem.

### 1.3 Målformulering

Målet med examensarbetet är att undersöka hur en uppsättning mikrofoner kan placeras/dimensioneras för att skapa fördelarna med en close talk-mikrofon från ett givet avstånd. De olika lösningarna ska utvärderas med avseende på prestandan och kostnaden. Slutmålet är att bygga en prototyp av en eller flera av lösningarna. Prototypen ska vara utformad för att kunna placeras i anslutning till/tillsammans med en pekskärm.

### 1.4 Problemformulering

Nedan formuleras problemställningar som är tänkta att besvaras under examensarbetet. "Bäst resultat" definieras här som bäst ljudåtergivning i termer av minst ljudförvrängning och brus, vilket valideras med PESQ-, STOI- och MOS-värde. Dessa värde beskrivs i avsnitt 2.6.1 och 2.6.2.

- Vilket antal mikrofoner ger bäst resultat?
- Vilken placering av mikrofonuppsättningen ger bäst resultat?
- Vilken utformning är lämpligast med tanke på att prototypen ska kunna placeras vid eller integreras med en pekskärm?
- Hur förhåller sig lösningarna till varandra kostnadsmässigt?

### 1.5 Motivering av examensarbetet

Valet av examensarbetet var motiverat av studenternas önskan av att få utveckla en produkt. Det var av stor vikt för studenterna att examensarbetet skulle fokusera på utveckling av hårdvara, även om en viss del programmering förekom i examensarbetet. Företaget ser examensarbetet som en teknisk studie.

### 1.6 Avgränsningar

Lösningarna ska vara avsedda för användning inom maximalt 1 meters avstånd och i kontorsmiljö. Lösningarnas arkitektur ska vara anpassad för att rymmas i/tillsammans med en pekskärm. Eliminering av ljudreflektioner från pekskärmen ingår inte i arbetet. Fokus ligger på utveckling av hårdvaran, och inte algoritmen för beamforming.

## 2 Teknisk bakgrund

### 2.1 Mikrofoner

Sensorer utformade för att fånga upp ljud inom hörbart intervall kallas generellt mikrofoner. I grunden är en mikrofon en trycktransduktor som är anpassad för att fånga upp ett stort spektrum av ljudvågor. Olika mikrofoner skiljer sig ifråga om exempelvis känslighet, riktningskaraktistik, akustisk bandbredd, dynamik och storlek. Olika användningsområden kräver olika lösningar.

Akustiska vågor är ett mekaniskt tryck, därför har mikrofoner samma basstruktur som en trycksensor. De innehåller ett rörligt membran och en transduktor som översätter membranets rörelser till elektriska signaler (Fraden, 2016). För detta används elektronikkomponenter med linjär karaktistik (Håkansson & Kjellander, 1991).

#### 2.1.1 Uppbyggnad

Det finns ett antal olika typer av mikrofoner, som skiljer sig ifråga om uppbyggnad och därmed egenskaper. Nedan beskrivs och förklaras de vanligast förekommande typerna.

##### Kolkornsmikrofon

När membranet vibrerar på grund av ljudvågen, ändras resistiviteten i materialet. Genom att koppla in det resistiva materialet i en elektrisk krets kan resistansändringen mätas. Transduktorn består av ett halvledande pulver, ofta grafit, vars bulkresistivitet är känsligt för tryck (Fraden, 2016). Svagheten hos dessa mikrofoner är brus, så kallad "carbon hiss", som inte kan elimineras utan bara reduceras genom noggrann förbehandling av pulvret (Calvert, 2003).

##### Kondensatormikrofon

Kondensatormikrofonen består av två plattor, över vilka det kopplas en spänning. En av plattorna är ett tunt plattformat membran som känner av signalen. Membranet är upphängt framför en annan metallplatta. När membranet utsätts för variationer i lufttrycket, sker en förändring i kapacitansen. Detta ger upphov till en laddningsförändring och kan i sin tur omvandlas till en spänning (Strong, 2002).

Kondensatormikrofoner är precisa och känsliga, dock dyra. Dessa mikrofoner kräver en matningsspänning mellan 10-200V (Fraden, 2016).

#### Elektretmikrofon

Elektretmikrofonen är nära besläktad med kondensatormikrofonen, med skillnaden att den förstnämnda genererar ett eget elektriskt fält och därmed har mikrofonkapseln inget behov av en extern spänningskälla. Elektretmaterial har mycket gemensamt med piezoelektriskt och pyroelektriskt material. En elektret är ett permanent polariserat dielektriskt material och kan betraktas som ekvivalent till en permanentmagnet. I mikrofoner används exempelvis tunna filmer av Teflon och Mylar. En elektretmikrofon är en elektrostatisk transduktor som består av ett metalliserat elektretmembran och en bakplatta, separerade med ett luftgap. Dessa mikrofoner är små, billiga, hållbara och användbara vid höga frekvenser (Fraden, 2016; Strong, 2002). Det tål att påpekas att även om mikrofonkapseln inte kräver matningsspänning så behöver de nödvändiga kringliggande komponenterna en spänningskälla.

#### Optisk mikrofon

Mikrofonens tillämpning är extrema miljöer, såsom raket- och jetmotorer. Mikrofonen är tålig, robust och kan användas vid vibrationer och höga temperaturer (Fraden, 2016). Dessa mikrofoner beskrivs inte vidare i detta arbete eftersom de saknar tillämpning i sammanhanget.

#### Piezoelektrisk mikrofon

I en sådan mikrofon används egenskaperna hos ett piezoelektriskt material, exempelvis kvartskristall, för att omvandla vibrationer till en elektrisk signal. De piezoelektriska egenskaperna innebär att det bildas en laddningsförskjutning som är proportionell mot kraften som påverkar kristallen. Mikrofonen har relativt låg känslighet och är relativt billig (Brandt, 2000). Den utmärker sig jämfört med andra mikrofoner genom att kunna placeras i olika medium (Fraden, 2016). Kristallmikrofoner användes förr mycket i enkla sammanhang eftersom de var billiga, dock förknippade med dålig ljudkvalitet (Studiefrämjandet, n.d.).

#### Dynamisk mikrofon

En dynamisk mikrofon använder elektromagnetisk induktion. Membranet sitter framför en spole, som hänger mellan de två polerna på en magnet. Membranets rörelser resulterar i att spolen rör sig och

genererar en svag elektrisk ström. Dynamiska mikrofoner är robusta och fukttåliga (Fraden, 2016). Dynamiska mikrofoner har nästan alltid ett lägre pris än kondensatormikrofoner (Studiefrämjandet, n.d.).

### Bandmikrofon

Principen för en bandmikrofon liknar densamma för den dynamiska mikrofonen. Skillnaden är att membranet, som hänger mellan de två polerna på en magnet, är av aluminium och inte av plast eller Mylar. Bandmikrofonens nackdelar är att den är dyrare och mindre robust än exempelvis kondensatormikrofoner. Bandmikrofonen utmärker sig genom att dämpa de höga frekvenser (Strong, 2002).

### MEMS-mikrofon

MEMS är en förkortning för Micro-Electro-Mechanical Systems. En sådan enhet innehåller både mekaniska och elektriska komponent. De flesta enheter skapade med MEMS-teknologi är i storleksordningen mikrometer (Fraden, 2016). MEMS-mikrofoner kan beskrivas som små mikrofoner med hög SNR, låg energikonsumtion och hög känslighet (Morcelli & Widder, 2014). Dessa är därför lämpliga för ljudapplikationer där liten mikrofonstorlek efterfrågas. Dessa mikrofoner är relativt billiga och kan erbjuda hög ljudkvalitet. Beamforming och brusreducering är några av tillämpningarna för dessa enheter. Därför kan de användas i exempelvis mobiltelefoner och bärbara datorer (STMicroelectronics, 2019).

### Sammanställning av olika mikrofontyper

Sammanfattningsvis finns det ett stort antal olika typer av mikrofoner. Dessa skiljer sig åt, och det är viktigt att ha mikrofonernas för- och nackdelar i åtanke vid val av mikrofontyp. Olika tillämpningar ställer olika krav på mikrofonerna. En översikt av de ovan beskrivna mikrofontyperna presenteras i Tabell I.

TABELL I. EN KORT SAMMANFATTNING ÖVER FÖR- OCH NACKDELAR HOS DE OLIKA MIKROFONTYPERNA.

| Mikrofontyp    | Fördelar                                                       | Nackdelar                        |
|----------------|----------------------------------------------------------------|----------------------------------|
| Kolkorn        |                                                                | Brus                             |
| Kondensator    | Precis och känslig                                             | Dyr, kräver hög matningsspänning |
| Elektret       | Liten, billig, hållbar och användbar vid höga frekvenser       |                                  |
| Piezoelektrisk | Kan användas i olika medium.<br>Billig                         | Låg känslighet                   |
| Dynamisk       | Robust, tålig. Billigare än kondensatortyp.                    |                                  |
| Band           |                                                                | Dyr och ömtålig                  |
| MEMS           | Små, billiga, hög SNR, låg energikonsumtion och hög känslighet |                                  |

### 2.1.2 Riktningsskarakteristik

Mikrofoner skiljer sig åt gällande riktningsskänslighet. Mikrofonens riktningsskarakteristik beskriver sättet på vilket mikrofonen fångar upp ljud, och kan delas upp i *rundupptagande* (omnidirektionella), *riktade* (unidirektionella) och *åttakopplade* (kallas också tvåriktade) mikrofoner (Strong, 2002).

### Polärddiagram

För att illustrera riktningsskarakteristiken används ofta polärddiagram, som visar förstärkning i förhållande till riktningens vinkel av infallande ljud. Mikrofontypen benämns efter formen av förstärkningsmönstret (Kjellson, 2014). Skisser på polärddiagram återfinns i Appendix A. Från polärddiagram för mikrofonuppsättningar framgår även huvudloben och eventuellt sidolober. Huvudloben pekar ut riktning i vilken mikrofonen är mest känslig, och lobens bredd beskriver mikrofonens upplösning. Storleken på sidoloberna illustrerar mikrofonuppsättningens oförmåga att undertrycka det omgivande bruset (Cigada, Ripamonti & Vanali, 2007).

### Rundupptagande mikrofoner

Omnidirektionella mikrofoner svarar på ljudtrycket från alla riktningar. Trycket inuti mikrofonen motsvarar omgivningens tryck. När en ljudvåg når mikrofonen, uppstår en tryckförändring på utsidan av membranet (Håkansson & Kjellander, 1991).

### Riktade mikrofoner

Unidirektionella mikrofoner svarar på ljudtrycket från en bestämd riktning. De direktionella egenskaperna hos mikrofoner uppnås mekaniskt. Detta möjliggörs genom att exempelvis ha två olika vägar till membranet: den direkta vägen och vägen genom en så kallad labyrinth, vars syfte är att fördröja ljudvågens väg till membranet (Håkansson & Kjellander, 1991).

### Åttakopplade mikrofoner

Dessa mikrofoner skiljer sig från rundupptagande mikrofoner genom att inte ta upp ljud från sidorna, utan enbart framifrån och bakifrån (Strong, 2002).

### Proximity effect (närhetseffekten)

Om riktade mikrofoner placeras nära ljudkällan, kan det uppstå så kallat proximity effect, som kan beskrivas som en bashöjning. Ju närmare ljudkällan mikrofonen befinner sig, desto mer bas fångar mikrofonen upp. Rundupptagande och åttakopplade mikrofoner visar dock ingen närhetseffekt. Även många av de riktade kondensatormikrofonerna har en inbyggd basavskärmning i syfte att undvika en sådan bashöjning (Strong, 2002).

### 2.1.3 Slutsats utifrån litteraturstudien

Bedömningen blev att elektretmikrofoner passade examensarbetet bäst. Anledningen till slutsatsen var att i tillämpningen eftersöktes mikrofoner med bra ljudåtergivning, vilket exkluderat kolkorns- och piezoelektriska mikrofoner. Önskemål om ett lågt pris på mikrofonerna exkluderade bandmikrofoner, som dessutom är ömtåliga jämfört med kondensatormikrofoner. De sistnämnda benämns också som dyra i litteraturen och kräver även högre matningsspänning. Det initialt tilltänkta utvärderingskortet var inte anpassat för MEMS-mikrofoner hårdvarumässigt. Sammanfattningsvis togs beslutet i samråd med handledaren på Axis Communications AB att enbart använda elektretmikrofoner i examensarbetet.

## 2.2 Mikrofonuppsättning

Det finns ett antal geometriska sätt att utforma konstruktionen av mikrofonuppsättningen. Dessa sätt har olika för- och nackdelar och måste därför väljas efter den tilltänkta tillämpningen. Mikrofonerna kan placeras en-, två- eller tredimensionellt.

### 2.2.1 Dimensionell placering av mikrofoner

#### Endimensionell uppsättning

En linjär uppsättning innebär att alla mikrofoner placeras på en rad efter varandra, antingen med samma avstånd mellan mikrofonerna (*uniform*) eller olika avstånd (*icke-uniform*). Icke-uniformitet minskar förekomst av grating lobes, det vill säga sidolober (Hald & Christensen, 2002), dock visas ingen påverkan i talområdet (Nisula & Krill 2015).

#### Tvådimensionell uppsättning

En sådan uppsättning innebär att mikrofonerna placeras längs figurens perimeter och/eller fyller upp arean inuti figuren. En tvådimensionell uppsättning kan uppnås genom att placera mikrofoner rektangulärt, cirkulärt, i multiring eller i en så kallad Spiralarmsuppsättning. Rektangulär uppsättning innebär att mikrofonerna placeras i form av en rektangel. I cirkulär uppsättning placeras mikrofonerna i form av en cirkel. Multiring-uppsättning betyder att mikrofonerna placeras längs flera cirklar. Med Spiralarmsuppsättning menas geometri där mikrofoner placeras längs flera cirklar i ett speciellt mönster.

### Tredimensionell uppsättning

I en sådan uppsättning placeras mikrofoner på en sfärisk yta (Lai, et al., 2017).

### 2.2.2 Mikrofonuppsättningen utifrån konsolens geometri

Konsolens utformning ställer krav på mikrofonuppsättningens egenskaper. Eftersom konsolens ovansida är rektangulär, undersöktes mikrofonuppsättningen med linjär geometri.

### 2.2.3 Avstånd mellan enskilda mikrofoner

Avstånd mellan enskilda element i en mikrofonuppsättning är väsentlig vid konstruktionen. Avståndet mellan mikrofonerna kommer att definiera den lägsta frekvensen (Ögren, 1996). För bredbandiga beamformer väljs  $\lambda$  som signalen med kortast våglängd, vilket motsvarar den högsta frekvensen. Detta val garanterar att spatial aliasing, inte ska förekomma upp till den valda frekvensen (Lai, et al., 2017). Spatial aliasing innebär att huvudloben blir smalare och antalet sidolober ökar, vilket medför att det blir omöjligt att bestämma riktningen för den inkommande signalen (Postema & Davidsson, 2019). Våglängden ( $\lambda$ ) beror på ljudets hastighet ( $v$ ) i media och dess frekvens ( $f$ ) enligt

$$\lambda = \frac{v}{f} \quad (1)$$

(Lai, et al., 2017).

För ljudvågor i luften är ljudets hastighet  $v = 343,59$  m/s vid rumstemperatur och normalt atmosfärtryck. Ljudets hastighet beror också på media och temperatur (Fraden, 2016), men eftersom den sökta tillämpningen för mikrofonuppsättningen är kontorsmiljö, beskrivs dessa samband inte här utan det ovan angivna värdet på  $v$  förutsätts. I syfte att undvika spatial aliasing bör det minsta avståndet ( $d$ ) mellan sensorerna beräknas enligt

$$d = \frac{\lambda}{2} \quad (2)$$

(Lai, et al., 2017).

I projektet konstruerades mikrofonuppsättningen för att fånga upp mänskligt tal. Våglängden för det mänskliga talet anges ligga mellan 250 och 8000 Hz (Götahälsan, 2016). En övre gräns på 3400 Hz tillåter röstigenkänning av en person, vilket kallas telefonkvalitet (Freeman, 2005).

Med hjälp av formlerna (1) och (2) beräknades minsta avståndet mellan mikrofonerna, se Tabell II. I examensarbetet används avståndet 45 mm mellan mikrofonernas centrum.

TABELL II. AVSTÅND MELLAN MIKROFONERNA I EN UPPSÄTTNING BERÄKNAT PÅ FREKVENSEN.

| Frekvens, kHz | Våglängd $\lambda$ , mm | Avstånd $d$ mellan mikrofonerna, mm |
|---------------|-------------------------|-------------------------------------|
| 8             | 42,9                    | 21,5                                |
| 4             | 85,9                    | 42,9                                |
| 3,4           | 101,0                   | 50,5                                |

#### 2.2.4 Antal element

Skillnad i ljudupptagningen mellan mikrofonuppsättningarna med två, tre respektive fyra element undersöktes. Det förväntade resultatet var att mikrofonuppsättningen med två mikrofoner skulle ge sämre resultat jämfört med en uppsättning bestående av tre eller fyra mikrofoner, samtidigt som skillnaden mellan två sistnämnda skulle vara liten. Detta antagande baserades på resultat påvisade av Nisula och Krill (2015) samt Kjellson (2014).

#### 2.2.5 Placering

Skillnad i mikrofonuppsättningens egenskaper undersöktes utifrån placering på fronten eller ovansidan av konsolen. Litteraturstudien gav ingen indikation på om någon placering skulle vara fördelaktig.

### 2.2.6 Sammanfattning av de mikrofonuppsättningsalternativ som undersöktes i examensarbetet

Efter den genomförda litteraturstudien undersöktes skillnaden mellan följande varianter

- Mikrofonuppsättning med två, tre och fyra element
- Placering av mikrofonuppsättningen på fronten och på ovensidan av konsolen

Därutöver undersöktes även hur olika mikrofontyper påverkar mikrofonuppsättningarnas egenskaper.

## 2.3 Beamforming

Mikrofonuppsättningen ska erbjuda möjlighet till beamforming, utformning av algoritmerna för beamforming ligger dock utanför examensarbetet. I detta avsnitt beskrivs kortfattat funktion av beamforming.

Beamforming är ett sätt att mottaga signaler från en viss riktning. Beamforming fokuserar mikrofonuppsättningen i en viss riktning, därifrån den önskade signalen kommer och samtidigt dämpar störningar från alla andra riktningar.

### 2.3.1 Konventionella beamformers

Den konventionella (klassiska) beamformingen innehåller tekniker såsom delay-and-sum, phase-shift, subband och filter-and-sum. Det gemensamma för teknikerna är att parametrar och viktningen är oberoende av indatan. Viktningen är anpassad efter störningarna i omgivningen (MathWorks, 2019).

#### Delay-and-sum beamformer

Beamforming av det enklaste slaget kallas delay-and-sum beamforming. Principen är att de lika viktade sensorerna mottar signal, och dessa signaler fördröjs och adderas, vilket resulterar i att den önskade signalen adderas konstruktivt, medan störningarna adderas destruktivt. Genom att justera viktningen i delay-and-sum beamformer kan man forma sidoloberna och strålen (beam). Denna teknik fungerar väl för signaler med smal bandbredd, men är inte så lämpliga i exempelvis talsammanhang. Där behövs olika viktning av frekvenserna (Benesty, et al., 2007).

### Filter-and-sum beamformer

Denna metod används när man är intresserad av mer flexibel lösning som tillåter kontroll av strålmönstret som funktion av frekvensen. Jämfört med delay-and-sum beamformer ersätts skalärviktning av sensorerna med en uppsättning av filter (Benesty, et al., 2007).

### 2.3.2 Adaptiva beamformers

Till skillnad från klassiska beamformers är de adaptiva beroende av indata. Minimum Variance Distortionless Response (MVDR) beamformer är ett exempel på en adaptiv beamformer. Om bruset och den önskade signalen saknar korrelation, är variansen hos den inspelade signalen summan av avvikelserna mellan den önskade signalen och bruset. MVDR-lösningen arbetar med att minimera denna summa och på det viset minska bruset. Det finns ett antal fördelar med MVDR, exempelvis högre spatialupplösning än för konventionell beamformer, samt mindre och mjukare sidolober. Nackdelarna, jämfört med klassiska beamformers, är högre beräkningskostnad och känslighet för fel (MathWorks, 2019).

## 2.4 Verktyg för in- och uppspelning av ljud

För ljud in- och uppspelning användes ett ljudkort och två mixerbord. RME Fireface 802 användes för inspelning av ljud från mikrofonuppsättningen, Behringer XENYX X1222USB användes för uppspelning av tal och Yamaha MG12X för uppspelning av brus.

### 2.4.1 Ljudkort för inspelning från mikrofonuppsättningen

RME Fireface 802.

Ljudkortet valdes baserat på rekommendation av handledaren på företaget. Ljudkortet har fyra mikrofoningångar som användes vid inspelningen. USB-kontakt användes för kommunikation mellan ljudkortet och datorn.

Gränssnittet som användes för att hantera ljudkortet är TotalMix FX. Det finns ett stort antal funktioner, det som användes var dock enbart styrningen av matningen till svanhalsmikrofonen (RME, n.d.).

## 2.4.2 Mixerbord för uppspelning av tal

Behringer XENYX X1222USB

Mixerbordet användes för taluppspelning. För detta användes en AUX-utgång från datorn till mixerbordet, och från mixerbordet till högtalaren (Genelec 8020C Studio Monitor) användes en XLR-kabel (Behringer, 2019).

## 2.4.3 Mixerbord för uppspelning av vitt brus

Yamaha MGP12X MIXING CONSOLE

Mixerbordet användes för brusuppspelning. För detta användes en AUX-utgång från datorn till mixerbordet, och från mixerbordet till högtalarna (Genelec 1037C Studio Monitor) användes XLR-kablar (Yamaha, n.d.).

## 2.4.4 Decibelmätare

För att mäta ljudnivån inför testinspelningarna användes en handhållen decibelmätare av typen Brüel & Kjær Type 2250-S. Mätverktyget är av klass 1, vilket normalt används för precisionsmätningar (Brüel & Kjær, n.d.; RISE, 2017).

Decibelmätarens frekvensvägningsfilter är ett A-filter, vilket innebär att filtret efterliknar örats känslighet vid svaga ljud. Den uppmätta ljudnivån är således A-vägd (Decibelmätare, 2019).

Inför mätningen kalibrerades decibelmätaren med för detta avsedd ljudkalibrator. Kalibreringen påvisade en avvikelse på 0,1 dB, vilket det justerades för under efterföljande mätningar.

## 2.4.5 Högtalare

### Genelec 8020C Studio Monitor

Denna högtalare användes för uppspelning av tal. Högtalaren är designad för att användas i en stor variation av miljöer och är lämplig vid närfältsapplikationer (Genelec, n.d.b).

### Genelec 1037C Studio Monitor

Två högtalare av denna sort användes för att spela upp vitt brus. Enligt tillverkaren har högtalarna mycket hög frekvensrespons och är perfekta för stereouppspelning, vilket användes under examensarbetet (Genelec, n.d.a).

### GRAS 44AA

Denna högtalare, benämnd i arbetet som munsimulator, användes för uppspelning av tal i det ytterligare testet. Högtalaren är designad för att efterlikna den mänskliga munnens dynamik (GRAS, n.d.a).

## 2.4.6 Hörlurar

### Shure SHR940

Dessa hörlurar användes vid subjektivt lyssnartest. Enligt tillverkaren erbjuder hörlurarna en mycket bra representation över hela audiospektrum med minimal förvrängning (Shure, 2019).

## 2.4.7 Mikrofonerna i examensarbetet

Två olika mikrofontyper användes under testinspelningarna. Båda typerna var omnidirektionella och av elektrettyp. Mikrofoner med diametern 6 mm benämns vidare i arbetet som mikrofoner av typ A, och typ B för mikrofoner med diametern 10mm. Mikrofonerna får inte beskrivas utförligare på grund av företagspolicyn. Fakta som tillåts beskrivas återfinns i Tabell III.

I det ytterligare testet användes en mätmikrofon av typen GRAS 26CA som är en precisionsmikrofon (GRAS, n.d.b)

TABELL III. BESKRIVNING AV MIKROFONER UTVALDA FÖR ATT ANVÄNDAS I EXAMENSARBETET.

| Typ | SNR vid 1 kHz | Parallelresistans vid 5V matningsspänning | Parallelresistans vid 3V matningsspänning | Seriekondensator |
|-----|---------------|-------------------------------------------|-------------------------------------------|------------------|
| A   | 60 dB         | 2,2 k $\Omega$                            | 1,3 k $\Omega$                            | 1 $\mu$ F        |
| B   | 80 dB         | 3,9 k $\Omega$                            | 2,2 k $\Omega$                            | 1 $\mu$ F        |

## 2.5 Digitala bearbetningsverktyg

I detta kapitel beskrivs verktyg som användes för bearbetning av data i ljudfiler.

### 2.5.1 MATLAB

MATLAB är ett beräkningsverktyg som är utvecklat av prof. Cleve Moler. MATLAB är ledande inom interaktiva numeriska program. Det finns ett stort antal tilläggsprogram för olika tillämpningar, och inkluderar grafik och simulering (Grimsberg, 2005).

I examensarbetet användes MATLAB i ett flertal syften, såsom tidskorrelering av ljudfiler, simulering av beamforming och uträkning av PESQ- och STOI värde.

### 2.5.2 Audacity

Audacity är ett verktyg för hantering av audiofiler, utvecklat av Audacity Team. I programmet finns det stöd för bland annat in- och uppspelning, återsampling, redigering av audiofiler samt användning av effekter och filter (Audacity, 2019).

I examensarbetet användes Audacity för hantering av ljudfiler. Det innebär ljudinspelning och uppspelning av vitt brus. Även återsampling och konvertering till .wav-filer skedde i Audacity. Verket användes även för redigering av ljudfilernas längd och innehåll baserat på korreleringsvärden framtagna i MATLAB.

## 2.6 Utvärderingsverktyg

För att utvärdera resultaten i den insamlade datan användes utvärderingsverktygen beskrivna i detta kapitel.

### 2.6.1 STOI

Short-time Objective Intelligibility (STOI) är en algoritm som beräknar taltydligheten genom att jämföra en referenssignal med en bearbetad signal. Metoden använder ljudsegment som är 386 ms långa.

Algoritmens resultat är ett värde mellan 0 och 1, där högre värde representerar bättre taltydlighet (Taal, Hendriks, Heusdens & Jensen, 2011).

### 2.6.2 PESQ

Perceptual Evaluation of Speech Quality (PESQ) är en standard innehållande metoder för utvärdering av talkvalité, utvecklad av The International Telecommunication Union (ITU). Standarden tillämpas ofta i telekommunikationssammanhang (International Telecommunication Union, 2018).

Algoritmen arbetar med två signaler, varav en är en referens- och en bearbetad signal. Resultatet är ett värde mellan -0,5 och 4,5, där ett lägre värde innebär större signalförvrängning. Metoden ger en indikation på ett förväntat MOS-värde. MOS (Mean Opinion Scores) är en metod för subjektiv bedömning av ljudkvalité, där en grupp människor lyssnar på en ljuduppspelning och poängsätter kvalitén mellan 1 (dålig) och 5 (utmärkt) (Loizou, 2013).

I examensarbetet betraktades främst värdet av Mean Opinion Score-Listening Quality Objective (MOS-LQO). Detta värde representerar en uppskattning av subjektiv "listening quality" (lyssningskvalité) som använder en objektiv värderingsteknik (International Telecommunication Union, 2007).

## 3 Metod

### 3.1 Arbetets faser

Arbetets gång innehöll följande faser: litteraturstudie, prototypframställning, testning, utvärdering, sammanställning av resultat samt rapportskrivande. Rapportskrivande och informationssökande pågick kontinuerligt under hela processen. Arbetsfasernas tidsåtgång illustreras i Gantt-schemat som framgår från Fig.1.

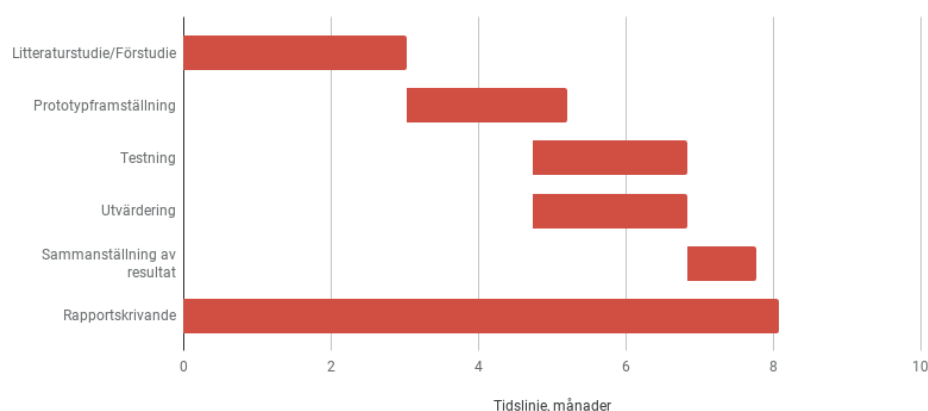


Fig. 1. Gantt-schema för tidsåtgång, där 0 representerar oktober 2018.

#### 3.1.1 Litteraturstudien.

Syftet med förstudien var att undersöka befintlig kunskap i ämnet genom att utföra en litteraturstudie. Litteratursökningen genomfördes i Lunds Universitets bibliotekskatalog LUBcat, nationell uppsatsdatabas UppSök samt artikeldatabas hos Audio Engineering Society. Resultaten från litteraturstudien är redovisade under kapitlet Teknisk bakgrund.

#### 3.1.2 Prototypframställning

I den här fasen arbetades arkitekturen för hårdvaran fram. Baserat på mikrofonernas datablad framställda för företaget beräknades dimensioner för komponenterna som behövdes för att konstruera

mikrofonuppsättningen. Elektronikkomponenterna och kablage införskaffades på företaget och kopplades samman. Först skedde kopplingen på ett kopplingsbord där oskärmade kablar användes. Efter utvärdering löddades komponenterna över på ett veroboard och skärmade kablar användes.

### 3.1.3 Testning

Under denna fas genomfördes ett antal tester: dels uppstartstester, för att se att utrustningen fungerade som det var tänkt, dels riktade tester som genererade resultat. Även subjektivt lyssnartest genomfördes under denna fas.

### 3.1.4 Utvärdering

Testerna genererade ett flertal resultat. En del av resultaten gav insikter om att vissa modifieringar behövde utföras (se avsnitt Prototypframställning).

### 3.1.5 Sammanställning

Sammanställningen av resultaten skedde genom införande av resultaten i tabeller och grafer. Dessa användes i nästa fas, rapportskrivningen, för att dra slutsatser och illustrera resultaten.

### 3.1.6 Rapportskrivning

Rapportskrivningen skedde kontinuerligt under arbetets gång i form av minnesanteckningar och utkast. Dessa sammanställdes i rapporten i arbetets slutskede.

## 3.2 Kommunikation

Alla arbetets faser skedde i nära samarbete med företaget där arbetet bedrevs. Dokumenten med anteckningarna delades via en plattform och samtal kring arbetet skedde kontinuerligt. Diskussioner har även bedrivits dagligen med handledaren på företaget. Samtal kring resultat och tillvägagångssätt var behovsstyrda. Halvtidspresentation av examensarbetet skedde under februari, där ett antal medarbetare och studenter närvarade. Värdefulla synpunkter togs emot och bearbetades, som en led i utvärderingen.

### 3.3 Val av metoder

Val av metoder och verktyg baserades på vetenskap och beprövad erfarenhet, med hänsyn till utrustningens tillgänglighet. Grundtanken var att använda befintliga komponenter/produkter på företaget, med tanke på eventuell vidareutveckling av produkten. Detta motiverades med att det är mer kostnadseffektivt för företaget att använda sig av beståndsdelar som redan beställs i större volymer, än att behöva förhandla inköp av en helt ny komponent/produkt. Dessutom har företagets medarbetare erfarenhet av produkten, vilket var värdefullt under examensarbetets gång. Sammanfattningsvis tillämpades väl beprövade system/verktyg (exempelvis Audacity, MATLAB) samt komponenter som redan fanns inköpta på företaget.

Valet av utvärderingsverktyg (PESQ och STOI) baserades dels på att metoderna är standardiserade och dels beprövade på företaget då dessa användes i andra examensarbete utförda på Axis Communications AB.

Eftersom de objektiva mätmetoderna (PESQ och STOI) beskrivs ge en grov indikation på upplevd subjektiv lyssnarkvalité, rekommenderas även ett genomförande av ett subjektivt kvalitetstest (Kondo, 2012). Därför utfördes även ett sådant enligt anvisningarna för testets utförande (Loizou, 2013).

Testuppsättningens spatiala placering rekommenderades av handledaren på företaget och baseras på en standardiserad metod (Brüel & Kjær, n.d).

### 3.4 Val av verktyg

Verktygen som användes var tillhandahållna av företaget och är väl beprövade av företagets medarbetare. Dessa verktyg används på Axis Communications AB vid testning i audiosammanhang och anses därför vara tillräckliga för att användas i examensarbetet.

### 3.5 Sätt att inhämta information

Informationen hämtades in kontinuerligt under arbetets gång. Ibland initierades informationssökandet med att handledaren på företaget föreslog användning av en viss metod eller komponent. En annan anledning till informationsinhämtning var avsaknad av en befintlig lösning på specifika problem, eller

nyuppkomna frågor som saknade svar under arbetets gång. För att fördjupa kunskapen om ämnet, undersöktes befintlig forskning och aktuell litteratur. Främst inhämtades informationen från vetenskapliga artiklar och böcker. Det förekom även sökningar på Internet efter relevant information för att bredda sökfältet. Denna information verifierades dock med vetenskapliga skrifter innan den godtogs.

## 3.6 Testuppställning och ljudinspelning

### 3.6.1 Ljudnivåer

Vid samtliga tester bestämdes ljudnivån på vitt brus till 50 dB peak-värde för att motsvara kontorsmiljön. Anledningen till den valda nivån är att den tilltänkta placeringen av konsolen är främst i kontorsmiljö. Talet spelades upp med ljudnivån 60 dB peak-värde, eftersom konversation anses hålla den nivån (Decibelmätare, 2019).

### 3.6.2 Val av ljud

#### Vitt brus

Vitt brus är brus vars frekvens- och energispektrum är konstant och oberoende av frekvensen. Signalenergin för en konstant bandbredd (centrerad kring en frekvens  $f_0$ ) ändras inte om  $f_0$  varierar. Namnet kommer från associationer till vitt ljus, där det finns lika mängd av alla färger (Carter & Mancini, 2009). Det vita bruset genererades med hjälp av verktyget Audacity.

#### Tal

För taluppspelningen användes så kallade Harvard sentences. Dessa meningar är en uppsättning av fraser som är standardiserade och används vid testning i exempelvis telekomsammanhang, där det finns ett behov av standardiserat och upprepningsbart tal. Det finns sammanlagt 72 listor á 10 meningar. Dessa fraser är fonetiskt balanserade och använder fonemer med samma frekvens som de förekommer i det engelska språket (IEEE Subcommittee on Subjective Measurements, 1969).

Meningarna i testet läses upp av en kvinna och en man och är inspelade i 16 kHz samplingsfrekvens. I Tabell IV redovisas de upplästa meningarna.

TABELL IV. HARVARD SENTENCES ANVÄNDA I TESTET.

| Uppläsarens kön | Uppläst mening                              |
|-----------------|---------------------------------------------|
| Kvinna          | It's easy to tell the depth of a well       |
| Man             | Kick the ball straight and follow through.  |
| Kvinna          | Glue the sheet to the dark blue background. |
| Man             | A pot of tea helps to pass the evening.     |

### 3.6.3 Testuppställning

Den första fasen i arbetet var att förbereda prototypen, konstruerad i kryssfaner, för mikrofonuppsättningen. Avstånden mellan mikrofonerna mättes upp och hålen för mikrofonerna av typ A borrades enligt Fig. 2 och Fig. 3.



Fig. 2. Konsolprototyp i kryssfaner, fotograferad framifrån.

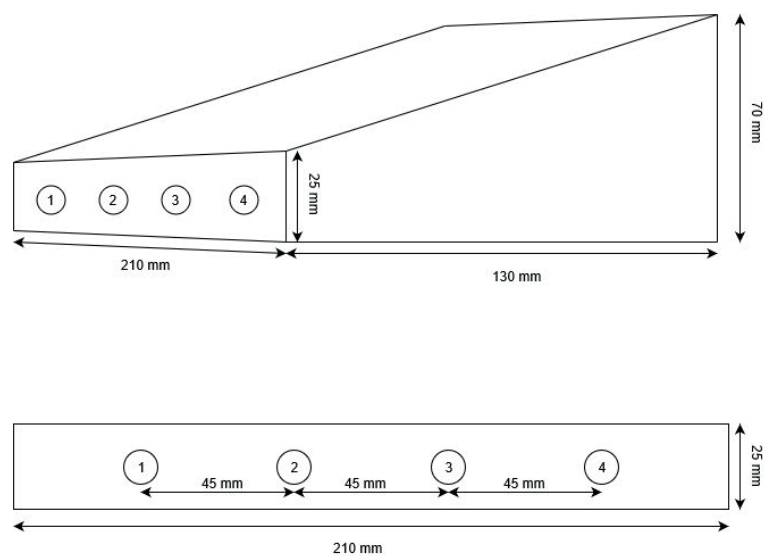


Fig. 3. Konsolprototypens mått.

Mikrofonerna kopplades på ett kopplingsbord enligt kopplingsschema redovisad i Fig. 4, och monterades därefter i konsolprototypen.

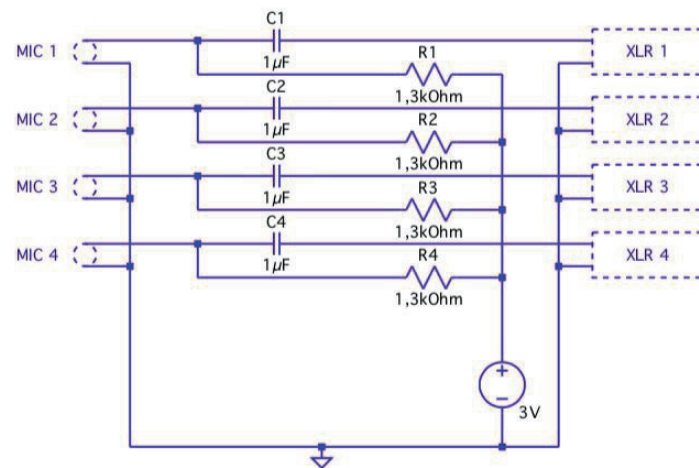


Fig. 4. Kopplingsschema för mikrofonuppsättningen med mikrofoner av typ A.

Kopplingsschema för mikrofonuppsättningen med mikrofoner av typ B skiljer sig från mikrofonuppsättningen med mikrofoner typ A gällande val av komponenter, se Fig. 5.

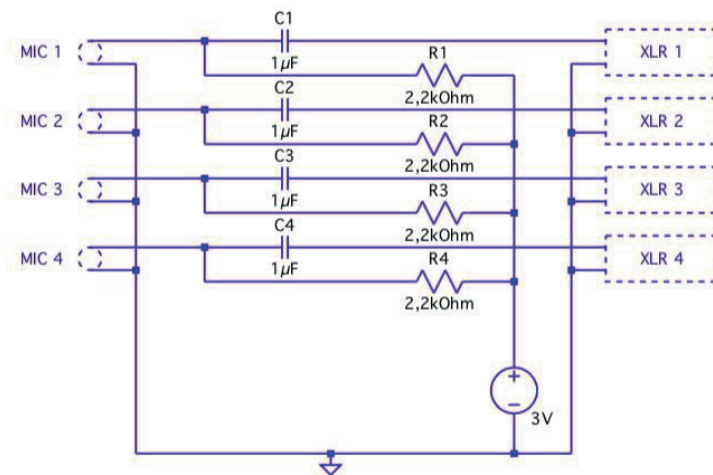


Fig. 5. Kopplingsschema för mikrofonuppsättningen med mikrofoner av typ B.

Därefter förbereddes testuppställningen i rummet, där väggarna var dämpade för att minska ljudreflektioner, enligt Fig. 6 och Fig. 7. Mixerborden för uppspelning kopplades till datorer och respektive högtalare. Mikrofonuppsättningen kopplades till ljudkortet för inspelning.

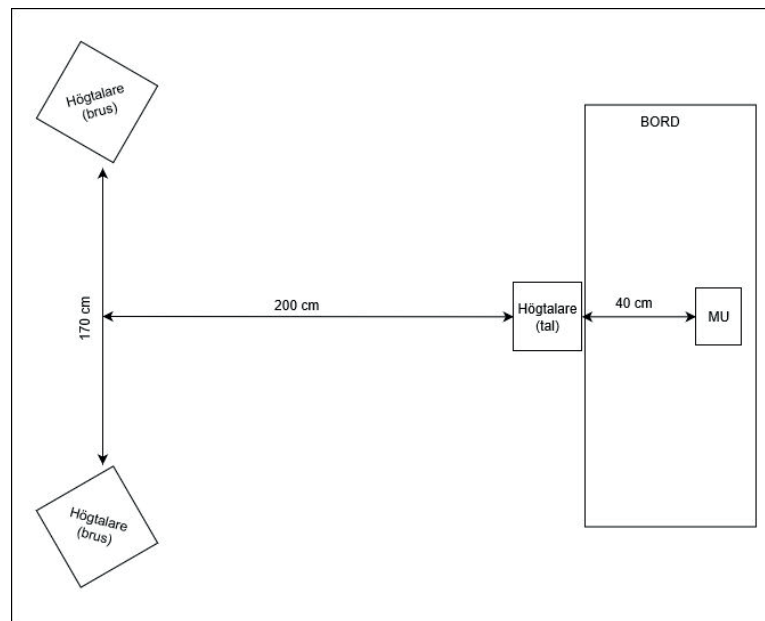


Fig. 6. Ritning över testuppställning, där MU är mikrofonuppsättningen i konsolen.



Fig. 7. Bild över uppsättningen av mixerbord och högtalare som spelade upp vitt brus. Konsolens placering på bordet och i förhållande till högtalaren framgår från Fig. 8 och Fig. 9.

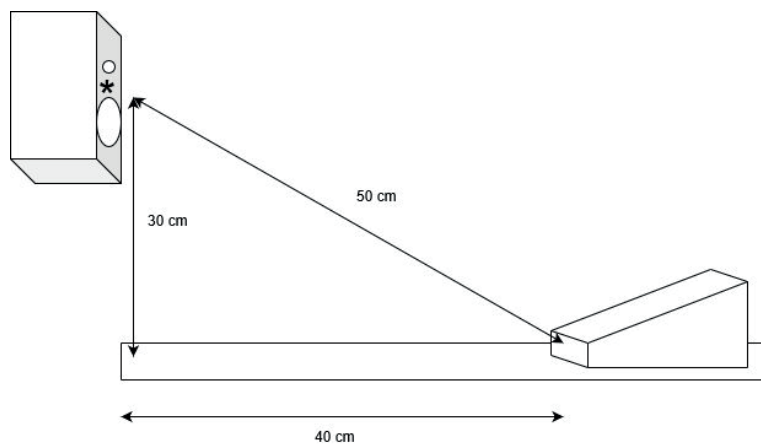


Fig. 8. Illustration av testuppställningen, där mikroupsättningens placering i förhållande till högtalarens akustiska centrum (\*) framgår (Genelec, 2013).



Fig. 9. Bild på testuppställningen. Denna uppställning är utförd enligt standardiserad metod (Bareham, 2018). Måtten framgår från Fig. 8.

För inspelningen via svanhalsmikrofonen användes samma placering av bordet och högtalaren. Avståndet mellan det akustiska centrumet hos högtalaren och svanhalsmikrofonen uppmättes till ca 10 cm, se Fig. 10.



Fig. 10. Bild på testuppställning för inspelning av svanhalsmikrofon.

För att säkerställa ljudnivåmätningarna utfördes normalisering (I) som bestod av följande steg:

1. Ljudfil med white noise (vitt brus) skapades och dess peak amplitud normaliserades till 0.0 dB i Audacity.
2. Ljudfilen med vitt brus spelades ut genom mixerbordet och volymen justerades till exakt 60.1 dB(A) peak-värde, eftersom decibelmätaren uppvisade avvikelse på -0.1 dB vid kalibrering.
3. Uppspelningen av vitt brus kopplades bort. Ljudfilen med klart tal togs in i Audacity och peak-amplituden normaliserades till 0.0 dB. På detta vis är det säkerställt att peak-värde i talet är 60 dB(A).
4. Det bortkopplade vita bruset kopplades till ett annat mixerbord och volymen justerades till 50 dB.

Efter normaliseringen utfördes inspelningen (II) via Audacity av tre varianter:

1. Endast tal
2. Tal och vitt brus
3. Endast vitt brus

Samplingsfrekvensen var 44.1 kHz. Inspelningen resulterade i fyra ljudfiler för varje variant, en för varje mikrofon i mikrofonuppsättningen. Den första mikrofonen utsågs att agera mätmikrofon.

Vidare kopplades den fjärde mikrofonen bort, så bara tre mikrofoner spelade in, och proceduren (II) upprepades.

Nästa steg var att koppla bort den tredje mikrofonen, och låta bara två mikrofoner spela in enligt (II).

Hela mikrofonuppsättningen kopplades bort och en inspelning (II) med en svanhalsmikrofon utfördes.

Då resultaten blev väldigt brusiga, kopplades mikrofonuppsättningen om, genom att använda veroboard istället för monteringsbord samt skärmade kablar istället för oskärmade. Efter att ha verifierat en förbättring via subjektivt lyssnande, genomfördes alla steg från med start från (I).

Därefter borrades hål på ovansidan av konsolen och de tidigare hålen fylldes med smältlim, se Fig. 11. Alla steg i inspelningen med start i (I) upprepades.



Fig. 11. Bild på konsolprototyp med mikrofon av typ A placerade på ovansidan.

Därefter borrades större hål för att de större mikrofonerna skulle kunna placeras i konsolprototypen och alla steg med start från (I) upprepades.

Hålen ovan limmades igen och det borrades nya, större hål på framsidan av konsolprototypen. Alla steg i inspelningen med start i (I) upprepades.

Detta resulterade i sammanlagt 111 inspelade ljudfiler.

### 3.6.4 Informationsbearbetning

Den initiala tanken var att använda ett utvärderingskort av typen AK7755-CHF, Nucleo 64 STM32L476 eller motsvarande. Ett flertal tester visade att det fanns begränsningar i korten ifråga om beamforming, och därmed användes MATLAB i detta syfte.

I examensarbetet användes MATLAB i ett flertal syfte, såsom tidskorrelering av ljudfiler, simulering av beamforming och uträkning av PESQ- och STOI värde. Koden till dessa funktioner återfinns i Appendix B.

Tillvägagångssättet för databearbetningen i MATLAB beskrivs nedan. I beskrivningen berörs bearbetningen av ljudfiler från en mikrofonuppsättning med fyra mikrofoner. Motsvarande process genomfördes med de andra uppsättningarna, där antalet mikrofoner var tre respektive två.

I ljudfilsbearbetningen användes följande ljudfiler:

- (a) Ljudfil med ljud av klart tal upptaget av första mikrofonen i mikrofonuppsättningen. Den användes som referensfil.
- (b) Ljudfiler med tal och brus, individuellt upptagna av fyra mikrofoner i mikrofonuppsättningen.

Bearbetningen i MATLAB följde stegen nedan:

1. Ljudfiler (b) användes som input till en beamformer-funktion, och resultatet blev en ljudfil med modifierat brusförvrängt tal (B).
2. Filer (a) och (B) användes som indata till korrelationsfunktionen, där utdata blev ett antal sampel, som filerna förflyttades i förhållande till varandra. Därefter klipptes filerna till samma längd. Förflyttningen/matchningen och klippningen utfördes i Audacity.
3. Dessa matchade filer användes som indata till PESQ- och STOI-funktionerna i MATLAB. PESQ-algoritmen krävde att samplingsfrekvensen på indatan 16 kHz, vilket innebar att ljudfilerna behövde återsamplas med den lägre frekvensen. Detta gjordes i Audacity.

4. Resultaten från steg 3 fördes in i Tabell V och VI och jämfördes med resultaten för övriga mikrofonuppsättningar och svanhalsmikrofonen.

För svanhalsmikrofonen utfördes alla steg utom den första, eftersom ljudet från en mikrofon inte kan beamformas.

Stegen för inspelningen och databearbetningen redovisas i Fig. 12 och Fig. 13.

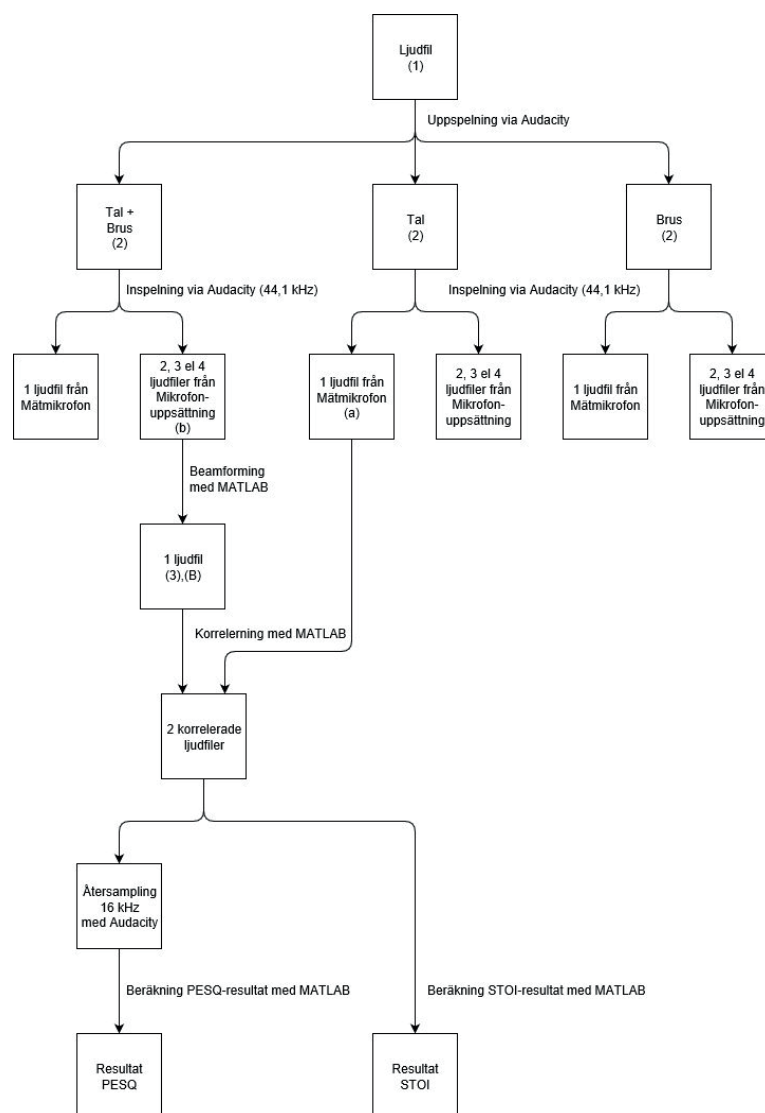


Fig. 12. Flödesschema för inspelning och databearbetning med en mikrofonupsättning. Förklaring av numreringen i exempelvis "Ljudfil (1)" återfinns i Fig. 14. Benämningarna (a), (b), (B) framgår ur processbeskrivningen i Kapitlet 3.6.4.

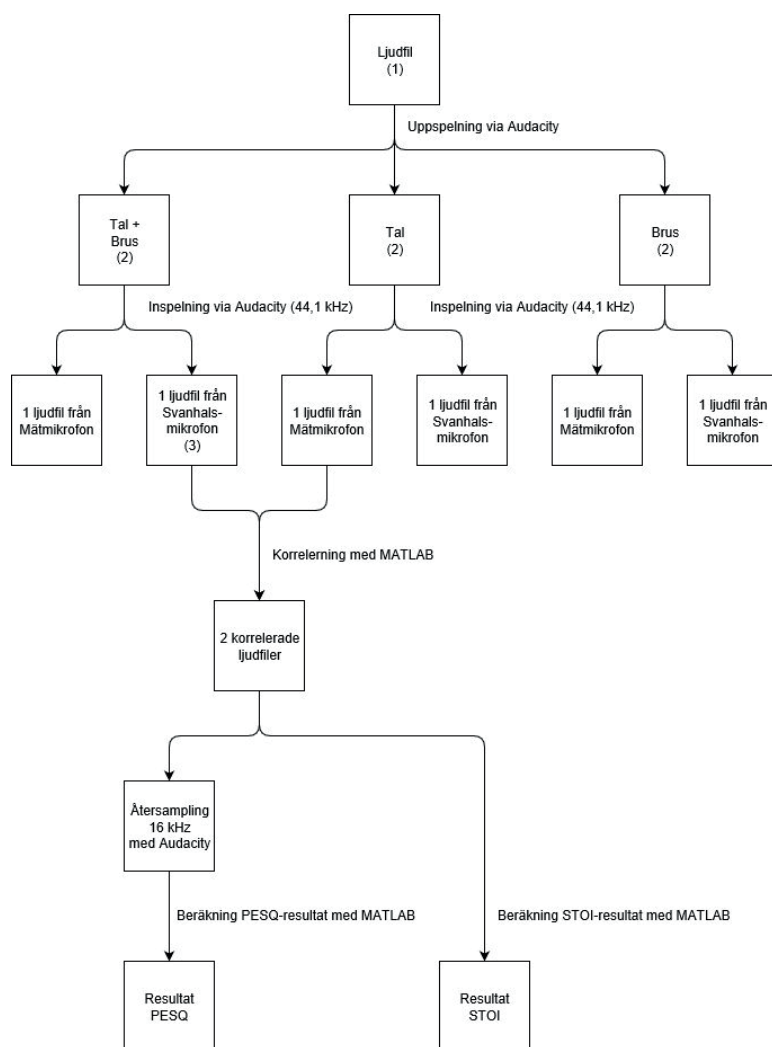


Fig. 13. Flödesschema för inspelning och databearbetning med en svanhalsmikrofon. Förklaring av numreringen i exempelvis "Ljudfil (1)" återfinns i Fig. 14. Benämningarna (a), (b), (B) framgår ur processbeskrivningen i Kapitlet 3.6.4.

I Fig.14 illustreras benämningen av olika ljudfiler, beroende på vilket steg i inspelningsprocessen det refereras till.

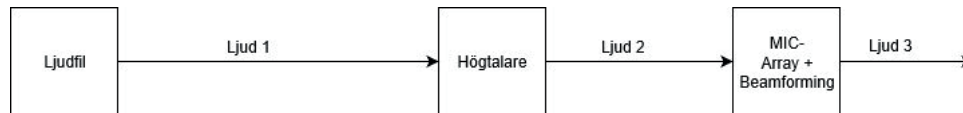


Fig. 14. Förklaring av ljudnumrering i Fig. 12, Fig. 13, Tabell V och Tabell VI.

### 3.6.5 Subjektivt lyssnartest

För att validera resultaten erhållna via MATLAB genomfördes ett subjektivt lyssnartest.

Rekommendationen är det minimala antalet ovana lyssnare är 20, maximala uppspelningstiden är 20 minuter samt att ljudet spelas upp i hörlurar (Loizou, 2013). Samtliga kriterier uppfyllades: 20 lyssnare fick lyssna på 13 ljudfiler à 11 s genom hörlurar. Testdeltagare fick lyssna på ljudfilerna och poängsätta varje ljudfil enligt MOS-skala. Formuläret som deltagarna fick fylla i återfinns i Appendix C. Innebörden av filernas numrering framgår ur Appendix D.

### 3.6.6 Ytterligare testinspelning med bäst lämpade mikrofonuppsättningar

Efter genomförda inspelningarna och mätningarna genomfördes utvärdering av resultaten. Utifrån utvärderingen valdes mikrofonuppsättningar och placeringen som gav bäst resultat. Dessa alternativ testades genom ytterligare en testinspelning. Denna inspelning skilde sig från föregående genom användning av en munsimulator för taluppspelning, och en specifik mätmikrofon som användes som referensmikrofon (GRAS 26CA). Dessutom centerades uppsättningen mot munsimulatoren. Det innebär att mikrofon nummer 2 och 3, mot tidigare 1 och 2, användes vid inspelningen med två mikrofoner samt att konsolen flyttades 22 mm sidledes för inspelningen med tre mikrofoner.

Munsimulators frekvenssvar testades före uppsättningen. För att det uppspelade talet skulle vara så nära originalet som möjligt, behövdes ett filter, som var inversen på munsimulators frekvenssvar. Detta filter återskapades manuellt i Audacity, baserat på de uppmätta värdena. Filterrepresentationen framgår i Appendix F. Testuppställning framgår från Fig. 15.

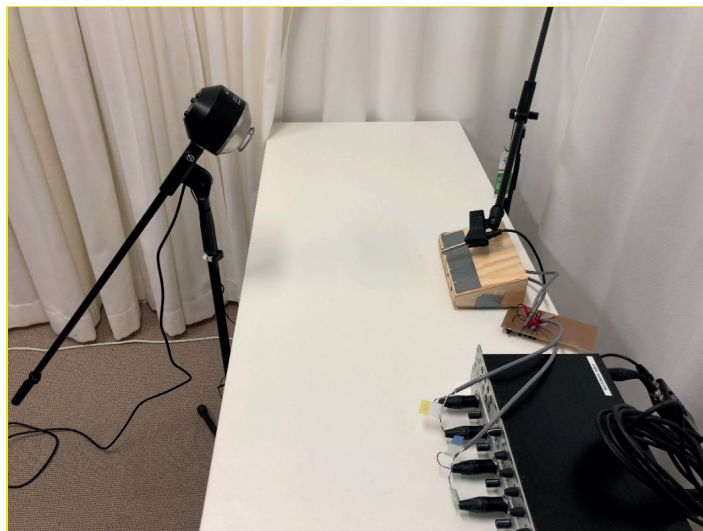


Fig. 15. Bild på testuppställning för uppspelning med munsimulator i det ytterligare testet.

### 3.7 Källkritik

De använda källorna består av vetenskapliga artiklar, genomförda examensarbeten, tryckt litteratur, presentationer från föreläsningar, hemsidor och slutligen hänvisningar till standarder och lagar. De vetenskapliga artiklarna är granskade före publicering, och anses därmed vara pålitliga. Examensarbeten, använda som referenser, har blivit granskade och godkända av respektive universitet och högskola, och anses därmed också vara pålitliga källor. Tryckt litteratur, som användes som referenser, är av sorten facklitteratur, och har blivit granskad av förlaget före publicering. Författare är experter inom sina respektive område, därmed anses källorna vara seriösa. Även presentationer från föreläsningar anses vara pålitliga källor, eftersom föreläsningarna hölls av områdesexperter. Den tekniska beskrivningen av använd utrustning kommer från tillverkarnas hemsida. Eftersom företagen inte får bedriva falsk marknadsföring, anses informationen vara tillförlitlig. Övriga hemsidor tillhör erkända organisationer och nätverk, och anses därmed kunna användas. Standarder är utfärdade av ITU (International Telecommunication Union), som är en autonom organisation, ansvarig för frågor som rör informations- och kommunikationsteknik. ITU är en av de äldsta mellanstatliga organisationerna, och standarder utfärdade av den anses vara pålitliga. Det var av yttersta vikt att hitta tillförlitliga källor för att bygga påståenden i examensarbetet på vetenskaplig grund, och samtliga källor som det refereras till i uppsatsen bedöms vara trovärdiga.

## 4 Analys

I detta avsnitt beskrivs olika val som gjordes under arbetets gång exempelvis val av mikrofoner, mikrofonuppsättning och programmeringsspråk. Vidare förklaras valens betydelse för arbetet. Därefter tas aspekterna kravspecifikation och kostnad upp och resultat från databearbetning presenteras i form av tabeller. Avslutningsvis beskrivs problem och deras lösningar.

### 4.1 Mikrofonval

Den gjorda bedömningen var att elektretmikrofoner passade examensarbetet bäst. Anledningen till slutsatsen var att i tillämpningen eftersöktes mikrofoner med bra ljudåtergivning, vilket exkluderar kolkorns-, piezoelektriska och dynamiska mikrofoner. Önskemål om ett lågt pris på mikrofonerna exkluderade bandmikrofoner, som dessutom är ömtåliga jämfört med kondensatormikrofoner. De sistnämnda benämns som dyra i litteraturen och kräver dessutom högre matningsspänning än elektretmikrofoner. Således användes elektretmikrofoner i examensarbetet.

### 4.2 Mikrofonuppsättningen

I examensarbetet ansågs enbart linjär uniform geometri vara lämplig. Anledningen till att de andra alternativen uteslöts var att den aktuella mikrofonuppsättningen kom att placeras i anslutning till en pekskärm i en rektangulär konsol. Det är olämpligt att placera några mikrofoner under skärmen såsom i multiring-uppsättning, Spiralarmsuppsättning och sfärisk uppsättning. Anledningen till att en sådan uppsättning är olämplig är att mikrofonerna under pekskärmen blir gömda, och kommer därmed inte ta upp någon signal av värde. Det är även olämpligt att använda cirkulär utformning då skärmen planeras vara rektangulär. Rektangulär geometri kommer inte tillföra något värde då talaren antas befinna sig framför talsystemet och inte förflytta sig nämnvärt. Eftersom tillämpningen avser användningen inom talområdet, användes uniform mikrofonuppsättning. Sammanfattningsvis undersöktes mikrofonuppsättningar av linjär uniform geometri.

### 4.3 Resultat från databearbetning

Nedan redovisas resultaten framtagna i MATLAB, se Tabell V-VI och Tabell VIII. Resultaten beräknade utifrån svaren i det subjektiva bedömningstestet återfinns att se i Tabell VII. Resultatvärdenas innebörd

förklaras i avsnitt 2.6.1 och 2.6.2. Mikrofontyperna framgår från Tabell III. Illustration över ljudnumreringarna (ljud 2, ljud 3) återfinns i Fig. 14.

TABELL V. STOI-VÄRDE, DÄR LJUD 2 ANVÄNDS SOM REFERENS OCH LJUD 3 SOM PROCESSAT LJUD.

| Placering | Antal<br>mikrofoner | Typ A  | Typ B  |
|-----------|---------------------|--------|--------|
| Ovansida  | 2                   | 0,6859 | 0,8916 |
| Ovansida  | 3                   | 0,7218 | 0,8989 |
| Ovansida  | 4                   | 0,7285 | 0,8906 |
| Front     | 2                   | 0,7463 | 0,9177 |
| Front     | 3                   | 0,7756 | 0,919  |
| Front     | 4                   | 0,7821 | 0,9039 |
| Svanhals  | 1                   | 0,7843 | 0,7843 |

TABELL VI. PESQ-VÄRDE, DÄR LJUD 2 ANVÄNDS SOM REFERENS OCH LJUD 3 SOM PROCESSAT LJUD.

| Placering | Antal<br>mikrofoner | Typ A | Typ B |
|-----------|---------------------|-------|-------|
| Ovansida  | 2                   | 2,4   | 2,794 |
| Ovansida  | 3                   | 2,508 | 2,964 |
| Ovansida  | 4                   | 2,507 | 2,922 |
| Front     | 2                   | 2,212 | 2,815 |
| Front     | 3                   | 2,414 | 3,014 |
| Front     | 4                   | 2,382 | 2,978 |
| Svanhals  | 1                   | 3,495 | 3,495 |

TABELL VII. RESULTAT AV MOS-TEST. INNEBÖRDEN AV LJUDFILERNAS NUMRERING FRAMGÅR FRÅN  
APPENDIX C. DE INDIVIDUELLA SVAREN FRAMGÅR UR APPENDIX E.

| Ljudfil nummer | Antal mikrofoner | Mikrofontyp | Placering | Resultat |
|----------------|------------------|-------------|-----------|----------|
| 1              | 1                | Svanhals    | Svanhals  | 2,8      |
| 2              | 2                | A           | Front     | 1,55     |
| 3              | 3                | A           | Front     | 1,95     |
| 4              | 4                | A           | Front     | 2,4      |
| 5              | 2                | A           | Ovansida  | 1,65     |
| 6              | 3                | A           | Ovansida  | 2        |
| 7              | 4                | A           | Ovansida  | 2,4      |
| 8              | 2                | B           | Front     | 3,35     |
| 9              | 3                | B           | Front     | 3,4      |
| 10             | 4                | B           | Front     | 3,45     |
| 11             | 2                | B           | Ovansida  | 3,35     |
| 12             | 3                | B           | Ovansida  | 3,4      |
| 13             | 4                | B           | Ovansida  | 3,3      |

TABELL VIII. RESULTAT FRÅN YTTERLIGARE TESTINSPELNING, BESKRIVET I AVSNITT 3.6.6, UTFÖRD MED MUNSIMULATOR OCH MÄTMIKROFON. MIKROFONUPPSÄTTNINGEN BESTOD AV MIKROFONER AV TYP B. PESQ- OCH STOI-VÄRDE, DÄR LJUD 2 ANVÄNDS SOM REFERENS OCH LJUD 3 SOM PROCESSAT LJUD.

| Placering | Antal mikrofoner | STOI   | PESQ  |
|-----------|------------------|--------|-------|
| Fronten   | 2                | 0,8292 | 2,414 |
| Fronten   | 3                | 0,8322 | 2,442 |
| Fronten   | 4                | 0,8408 | 2,429 |
| Svanhals  | 1                | 0,9308 | 2,895 |

## 4.4 Kravspecifikation

Med hänsyn till konsolens storlek fick mikrofonuppsättningens längd inte överstiga 210 mm, som förmodas bli konsolens totala bredd.

## 4.5 Kostnad

Den sammantagna kostnaden av produkten består inte enbart av komponenternas kostnad utan även av exempelvis monterings-, transport- och underhållskostnaden. Det går inte att ta fram priserna på enskilda element eftersom dessa varierar beroende på beställningsvolymen. Dessutom tillåter inte företaget att information om prisanbud på exempelvis mikrofoner lämnas ut. Därför redovisas resultaten i kostnadsjämförelser i andelar och inte i kronor. Det kan dock konstateras att det är nödvändigt att jämföra slutpriset på produkten med olika antal mikrofoner, och utvärdera dessa kostnader i förhållande till konsolens prestanda. Sammanfattningsvis måste en noggrann avvägning göras huruvida den ökade totalkostnaden är försvarbar med tanke på förbättringsgraden av ljudkvalitén.

## 4.6 Valens betydelse för resultatet

### 4.6.1 MATLAB vs Python

Parallellt med detta examensarbete pågick ett annat examensarbete på Axis Communications AB, som också handlade om beamforming. Fokuset i det arbetet var mjukvaruimplementering av beamforming och detta genomfördes i programmeringsspråket Python. I det här examensarbetet har dock MATLAB

använts eftersom det redan fanns både en färdig implementation av beamforming och ett stort bibliotek innehållande andra funktioner som eventuellt var användbara.

#### 4.6.2 Brusreducering

Initialt i arbetet användes kopplingsbord och oskärmade kablar. Detta kopplingssätt visade sig dock orsaka elektriskt brus och ändrades till koppling med veroboard och skärmade kablar. Syftet med detta var att inspelningarna före bearbetning skulle innehålla minimalt med komponentorsakade störningar.

### 4.7 Problem och lösningar

#### 4.7.1 Utvärderingskort

Initialt planerades användning av ett utvärderingskort för signalbearbetning. Detta stötte dock på flera problem under arbetets gång varvid ett byte av strategi krävdes.

#### 4.7.2 Avsaknad av manual till rätt version av utvärderingskort

Detta löstes genom att använda en manual till ett snarlikt utvärderingskort av en äldre version.

#### 4.7.3 Oklarheter kring val av programvara

Det fanns ingen programvara med rätta funktioner att tillgå till det befintliga utvärderingskortet. Då gjordes ett test med programvara för ett snarlikt kort. Det konstaterades att det inte fanns några lämpliga utvärderingskort med tillhörande mjukvara på företaget.

#### 4.7.4 Utvärderingskort kan bara ta in signal från två mikrofoner

Signalintaget bedömdes vara otillräckligt, varav byte av strategi skedde. I slutändan användes inte något utvärderingskort alls, utan funktionerna i MATLAB utnyttjades istället.

#### 4.7.5 Avsaknad av komponenter

Exempelvis resistorer fanns inte i den beräknade utformningen, så beslutet togs i samråd med sakkunnig på företaget att använda komponenter som var närmast det uträknade värdet. Även skärmade kablar fanns inte att tillgå i början av arbetet, utan införskaffades efter ett påvisat behov av sådana.

#### 4.7.6 Elektriskt brus

En första testinspelning påvisade mycket elektriskt brus. Detta löstes genom användning av veroboard istället för kopplingsbord samt skärmade kablar istället för oskärmade.

## 5 Resultat

I detta avsnitt besvaras de inledande frågeställningarna beskrivna i avsnitt 1.4. Med hjälp av resultaten redovisade i kapitel 4 ges förslag på lämpligt antal mikrofoner och placering av mikrofonuppsättningen. Även resultaten i det subjektiva lyssnartestet sammanfattas. Avslutningsvis diskuteras de föreslagna alternativen utifrån ett kostnadsperspektiv.

### 5.1 Antal mikrofoner

*Vilket antal mikrofoner ger bäst resultat?*

Utifrån Tabell V-VII kan det utläsas att tre eller fyra mikrofoner ger bättre ljudåtergivning än 2 mikrofoner. Det bekräftas även av resultaten i det ytterligare testet, se Tabell VIII.

### 5.2 Placering

*Vilken placering av mikrofonuppsättning ger bäst resultat?*

Utifrån Tabell V-VII kan det utläsas att mikrofoner av typ B placerade på fronten av konsolen ger bäst resultat i PESQ-, STOI- och MOS-värde. Det kan dock tilläggas att utifrån Tabell VII ger tre mikrofoner placerade på ovansidan samma resultat som tre mikrofoner placerade på fronten. Dessutom finns det en skillnad i resultaten avseende PESQ- och STOI-värde för mikrofontyp A, där Tabell V visar att frontplaceringen är mer fördelaktig än placeringen på ovansidan, medan Tabell VI visar motsatsen.

Företagets önskemål är att placeringen sker på fronten, vilket inte strider mot resultaten.

*Vilken utformning är lämpligast med tanke på att prototypen ska kunna placeras vid eller integreras med en pekskärm?*

I examensarbetet bedömdes enbart linjär geometri vara lämplig. Eftersom tillämpningen avser användningen inom talområdet, användes uniform mikrofonuppsättning.

### 5.3 MOS-test

Det subjektiva lyssnartestet bekräftade resultatet från MATLAB-beräkningarna. Även om poängen inte stämmer exakt numeriskt, så finns det en indikation på vilka mikrofonuppsättningar som gav bäst ljudåtergivning: nämligen tre eller fyra mikrofoner av typ B placerade på fronten av konsolen.

### 5.4 Kostnader

*Hur förhåller sig lösningarna till varandra kostnadsmässigt?*

Som tidigare nämnts, är det inte tillåtet att redovisa priset för enskilda komponenter. Det har dock konstaterats att mikrofoner av typ B är ungefär dubbelt så dyra som mikrofoner av typ A. Detta förhållande kan dock ändras beroende på beställningsvolym och aktuella avtal.

## 6 Slutsats

Utifrån Tabell V och Tabell VI kan det utläsas att tre mikrofoner av typ B placerade på fronten av konsolen ger bäst resultat i PESQ- och STOI-värde. Även fyra mikrofoner av samma typ och placering ger bra resultat, och ljudåtergivningen värderades som bäst i MOS-testet.

Det ytterligare testet påvisade snarlika resultat, med skillnaden att 3 mikrofoner gav högst PESQ-värde medan 4 mikrofoner gav högst STOI-värde. Detta gäller för mikrofoner av typ B placerade på fronten.

### 6.1 Reflektion över etiska aspekter

#### 6.1.1 Försiktighet vid fotografering

Före publiceringen av tagna bilder i examensrapporten fick dessa granskas av handledaren och avdelningschefen på företaget. Inga ännu ej släppta produkter tillverkade av Axis Communications AB fick förekomma på bilderna.

#### 6.1.2 Integritet

Produktlösningen framtagen i examensarbetet är inte tänkt att användas för inspelning av samtal. Det bör dock tas i beaktande att det finns en möjlighet att önskemål om inspelning kan förekomma. Inspelningen av samtal regleras i Brottsbalken, 4:e kapitlet, paragraf 9a: *"Den som ... olovligen medelst tekniskt hjälpmedel för återgivning av ljud i hemlighet avlyssnar eller upptager tal i enrum, samtal mellan andra eller förhandlingar vid sammanträde eller annan sammankomst, vartill allmänheten icke äger tillträde och som han själv icke deltar i eller som han obehörigen berett sig tillträde till, dömes för olovlig avlyssning till böter eller fängelse i högst två år."* (Brottsbalk, SFS 1962:700).

Det finns därför en etisk och juridisk aspekt att ta hänsyn till gällande användning av mikrofoner/ljudupptagningsutrustning. Det är viktigt att ta ställning till huruvida ljudet ska spelas in. Människor som befinner sig i närheten av konsolen kan vara ovetande om att deras samtal spelas in, om de inte uppmärksammas på det. Därför är det ett befogat ställningstagande om en sådan möjlighet ska finnas och i så fall hur man kan informera de berörda personerna om den.

### 6.1.3 Miljö

Elektronik innehåller en mängd olika kemikalier och tungmetaller som påverkar miljön under produktion, användning och skrotning. Kemikalierna avdunstar under elektronikanvändning och kan hamna i jord och vatten vid skrotning (Greenpeace, 2019).

När en ny produkt designas, finns det många egenskaper att ta hänsyn till. I examensarbetet har det framkommit att mikrofonuppsättning med tre mikrofoner ger bättre resultat än en konsol med två mikrofoner. Dock kommer miljöpåverkan vara större ju fler mikrofoner det används i konsolen, inte minst det eftersom det kommer krävas ett ökat antal av andra komponenter såsom resistorer, kondensatorer och kablage. Detta ökar både transportpåverkan och framställning av nya elektronikkomponenter. Även om miljövinsten med färre mikrofoner kan räknas som försumbar i en enskild konsol, så kommer den vara tydlig om man tillverkar ett stort antal av konsoler.

## 6.2 Framtida utvecklingsmöjligheter

### 6.2.1 Mikrofonuppsättningens geometri

I examensarbetet ansågs enbart linjär geometri vara lämplig. Rektangulär geometri kommer inte tillföra något värde då talaren antas befinna sig framför talsystemet och inte förflytta sig nämnvärt. Det är dock av intresse att i framtiden utvärdera den rektangulära geometrin, med tanke på möjlig vidareutveckling, t.ex. att systemet ska kunna följa en talare som förflyttar sig.

Eftersom tillämpningen avser användningen inom talområdet, användes uniform mikrofonuppsättning. Det kan även vara av intresse att undersöka den icke-uniforma uppsättningen om användningsområdet är annat än talsammanhang.

### 6.2.2 Avskärmning

En oavskärmd mikrofonuppsättning tar upp mycket brus bakifrån. En föreslagen åtgärd är avskärmning i form av en akustisk ljuddämpare (Shen, Chu, Yang & Wang, 2017).

### 6.2.3 Algoritmer för beamforming

Det här examensarbetet har fokuserat på utformningen av hårdvaran i mikrofonuppsättningen. Samtidigt har det konstaterats att resultatet kan påverkas av mjukvaruimplementationen för konsolen. Algoritmernas effektivitet i beamforming är beroende av förutsättningarna såsom konsolens tilltänkta upptagningsradie och önskad talkvalité. Det finns ett stort antal algoritmer för beamforming, och valet måste göras med hänsyn till förutsättningarna. Sammanfattningsvis finns det en utvecklingspotential och behov av ställningstagande gällande mjukvaruimplementationen.

### 6.2.4 Ytterligare efterbehandling av ljud

Under examensarbetets gång har fokuset legat på beamforming. Efterbehandling av filerna med det processade (beamformade) ljudet är möjlig, och gjordes i utforskande syfte. Fig. 16-19 erbjuder en möjlighet till jämförelse mellan ursprungligt ljud, inspelat ljud, beamformat ljud samt beamformat och ytterligare brusreducerat ljud. Utifrån spektrogrammen kan det se ut som om beamformingens förbättringseffekt är mindre än brusreduceringens. Det skulle kunna förklaras med att brusreduceringsalgoritmen är anpassad för att just eliminera brus och är därmed optimerad för detta syfte, vilket beamformingsalgoritmen inte är. Det tål dessutom att understrykas att det i examensarbetet inte gjorts försök att brusreducera den icke-beamformade ljudfilen, och därför kan jämförelse av algoritmernas effektivitet i fråga om brusreducering inte göras. Däremot kan det konstateras att kombination av algoritmer för beamforming och brusreducering kan förbättra resultatet.

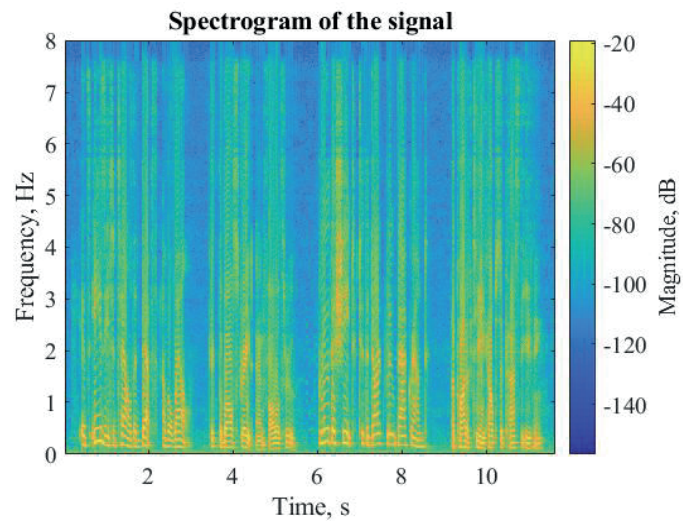


Fig. 16. Spektrogram över det ursprungliga ljudet med enbart tal.

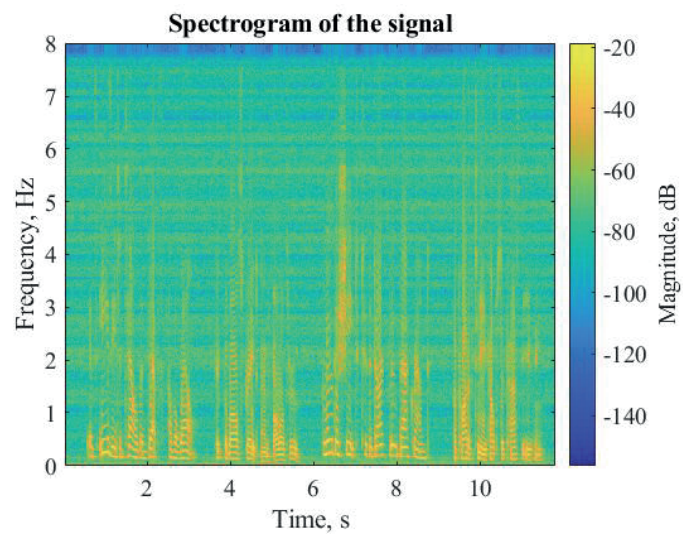


Fig. 17. Spektrogram över det inspelade ljudet, som bestod av tal och vitt brus. Spektrogrammet återspeglar ljudet inspelat med en mikrofon av typ B, placerad på fronten.

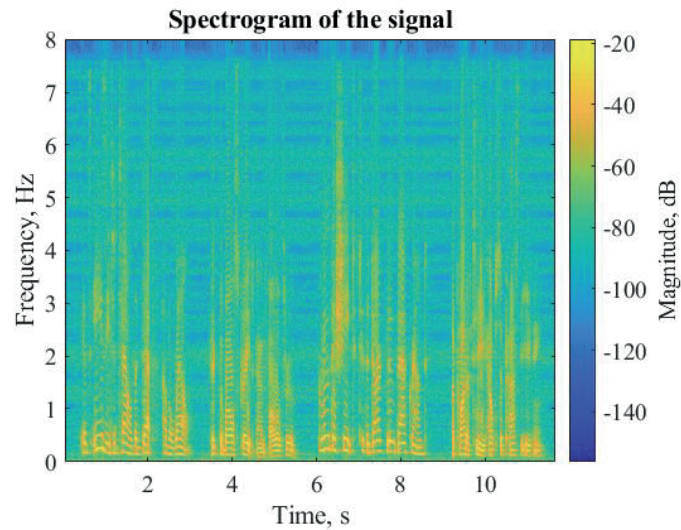


Fig. 18. Spektrogram över det inspelade ljudet, som bestod av tal och vitt brus, efter beamforming. Spektrogrammet återspeglar ljudet inspelat med tre mikrofoner av typ B, placerad på fronten.

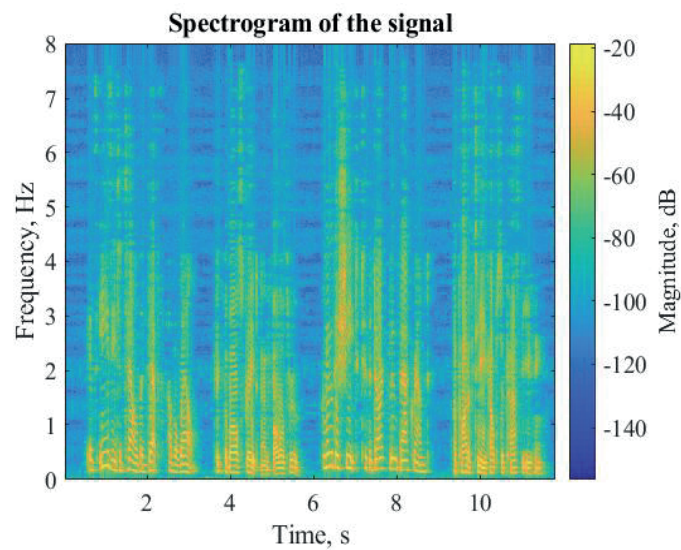


Fig. 19. Spektrogram över det inspelade ljudet, som bestod av tal och vitt brus, efter beamforming och brusreducering. Spektrogrammet återspeglar ljudet inspelat med tre mikrofoner av typ B, placerad på fronten.

### 6.2.5 Fler inspelningar

I examensarbetet utfördes ett begränsat antal inspelningar. Det hade varit av värde att utföra fler inspelningar, förslagsvis med andra tekniker och annan utrustning, för att verifiera de erhållna resultaten. Som exempel kan nämnas inspelning med toner, annan typ av brus och annorlunda placering av ljudkällorna.

Initialt i examensarbetet ändrades inte konsolplaceringen i förhållande till talljudkällan mellan inspelningarna med olika antal mikrofoner. I det ytterligare testet placerades munsimulatoren vinkelrät mot mikrofonuppsättningens centrum under alla inspelningar. I utforskande syfte gjordes även en inspelning med en icke-centrerad konsol. Avvikelsen för den icke-centrerade konsolen motsvarar  $2,9^\circ$ , som framgår i Fig. 20. I Tabell IX redovisas skillnaderna i STOI- och PESQ-värden för inspelningar med centrerad respektive icke-centrerad konsol utförda med tre mikrofoner av typ B placerade på fronten.

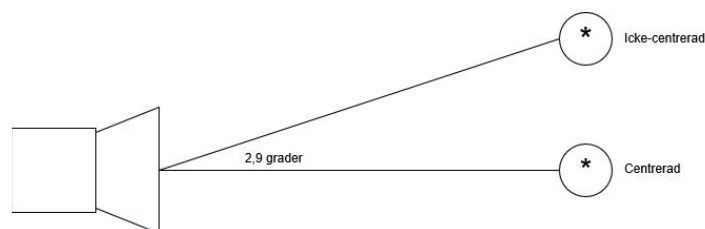


Fig. 20. Skiss över avvikelsen i grader mellan den centrerade och icke-centrerade konsolens mitt.

\* representerar mikrofonuppsättningens centrum.

TABELL IX. RESULTAT FRÅN INSPELNING MED CENTRERAD KONTRA ICKE-CENTRERAD KONSOL.

|                | STOI   | PESQ  |
|----------------|--------|-------|
| Icke-centrerad | 0,8277 | 2,442 |
| Centrerad      | 0,8322 | 2,442 |

Skillnaden mellan resultaten skulle kunna vara annorlunda beroende på beamformingsalgoritmen. Exempelvis beskrivs MVDR-beamformer vara känsligare för riktningsvinkeln än delay-and-sum-beamformer (Postema & Davidsson, 2019). Sammanfattningsvis är det av intresse att utföra fler inspelningar med andra uppställningar beroende av exempelvis valet av algoritmer.

### 6.2.6 Användning av MEMS-mikrofoner

I examensarbetet har det använts elektronikkomponenter som fanns tillgängliga på företaget. Beroende på önskad design av konsolen, hade man kunnat prova använda MEMS-mikrofoner.

### 6.2.7 Utvärderingskort

Initialt i arbetet undersöktes möjligheten att använda ett utvärderingskort i syfte att bearbeta ljudet i realtid. Detta visade sig vara problematiskt eftersom de befintliga korten på företaget inte uppfyllde kriterierna i form av tillgänglig mjukvara eller antal möjliga ljudingångar. Det är dock önskvärt att hitta ett utvärderingskort för tillämpningen, eftersom ljudet behöver bearbetas i realtid när mikrofonuppsättningen används i en talkonsol. Därför genomfördes sökningar efter utvärderingskort, anpassade för beamforming av tal, på hemsidor hos följande tillverkare/återförsäljare: Elfa Distrelec, Digi-Key, Farnell, RS Components, Texas Instruments, Analog Devices, Seeed Technology, Maxim Integrated, Nexperia, STMicroelectronics och Asahi Kasei Microdevices Corporation. Sökningarna har fokuserat på att hitta utvärderingskort anpassade för beamforming av ljud från fler än två mikrofoner. Önskemål om ett större antal mikrofoner än två har därmed begränsat sökresultatet. Det finns ett antal kort, utöver de nedan nämnda, som är anpassade för beamforming av ljud från just två mikrofoner.

Sökningarna resulterade i ett antal förslag för framtida undersökningar. Nedan beskrivna alternativ skulle kunna användas för beamforming av ljud från fler än två mikrofoner i en talkonsol. Inget av korten ter sig dock vara exakt anpassade efter mikrofonuppsättningen i examensarbetet, utan behöver anpassas mjuk- och/eller hårdvarumässigt. Då mjukvaran för dessa kort inte är öppet tillgänglig, är det svårt att precisera behovet av kodändringar. Hårdvaruanpassningar bör göras med hänsyn till mikrofonerna utvalda för tillämpningen. Sammanfattningsvis är det svårt att beskriva exakt vad som behöver ändras i varje alternativ utan att först precisera tillämpningen och få tillgång till koden. Därför beskrivs förslagen nedan enbart översiktligt och utan att diskutera nödvändiga anpassningar.

#### TMDSEVM5517

Denna utvärderingsmodul kan användas för att extrahera ett klart tal från en brusigt miljö med hjälp av algoritmer för exempelvis beamforming, där ljud upptaget av två till fyra mikrofoner används (Texas Instruments, 2019).

### MAX9860EVKIT

Detta utvärderingskit innehåller MAX9860-utvärderingskort och tillhörande mjukvara. Dessa tillåter inspelning från en mikrofonuppsättning med två eller fyra mikrofoner och användning av en beamformingsalgoritm (Maxim Integrated, 2019).

Utöver ovan nämnda alternativ finns det även ett antal lösningar, där mikrofonerna är förmonterade på kortet. Det innebär att mikrofonerna och geometrin på mikrofonuppsättningen skiljer sig från den i examensarbetet.

### ReSpeaker Mic Array v2.0

Detta kort kan applicera beamformingsalgoritmer på ljud upptaget med hjälp av cirkulär mikrofonuppsättning bestående av fyra förmonterade mikrofoner (Sceed Studio, 2019a).

### ReSpeaker Core v2.0

Detta kort använder ljud upptaget med hjälp av sex mikrofoner, förmonterade i en cirkulär mikrofonuppsättning, för beamforming (Sceed Studio, 2019b).

Alternativen ovan bör ses som förslag på utvärderingskort som kan användas i framtida arbete. Dessa förslag bör testas och anpassas efter tillämpningen. Sammanfattningsvis är det av värde att kunna testa mikrofonuppsättningen i realtid med hjälp av exempelvis de ovan nämnda utvärderingsmodulerna.

## 6.2.8 Sammanfattning av framtida utvecklingsmöjligheter

Vidareutveckling av produkten kan innebära utvecklingsmöjligheter och nya användningsområden för mikrofonuppsättningen. Exempelvis kan förändrad geometrisk form av konsolen innebära möjlighet till exempelvis cirkulär uppsättning. En annan variant är om det finns ett önskemål om att följa talaren som förflyttar sig: för detta skulle det krävas minst en L-formad uppsättning och en annan algoritm för beamforming. Det skulle kräva fler inspelningar och utvärderingar. Högre krav på ljudkvalité kan framtvinga behov av ytterligare brusreducering. Nya upphandlingar av komponenter kan erbjuda annorlunda mikrofontyper. Testning av mikrofonuppsättningen i realtid med hjälp av utvärderingsmoduler anpassade för syftet kan ge nya insikter. Sammanfattningsvis kommer utvecklingen av själva konsolen och företagets inköp leda till förändrade villkor för mikrofonuppsättning och därmed en eventuell förändring av denna.

## 7 Terminologi

**Beamforming** är ett sätt att mottaga signaler från en viss riktning. I detta arbete undersöks mottagande av en signal. Beamforming fokuserar mikrofonuppsättningen i en viss riktning, därifrån den önskade signalen kommer och dämpar samtidigt störningar från alla andra riktningar.

**MEMS-mikrofoner** (Micro-Electrical-Mechanical System) kallas också för ett mikrofonchip. Dessa kan beskrivas som små mikrofoner med hög SNR, låg energikonsumtion och hög känslighet.

**MOS** (Mean Opinion Scores) är en metod för subjektiv bedömning av ljudkvalité, där en grupp människor lyssnar på en ljuduppspelning och poängsätter kvalitén mellan 1 (dålig) och 5 (utmärkt).

**MOS-LQO** (Mean Opinion Score-Listening Quality Objective) innebär en uppskattning av subjektiv "listening quality" (lyssningskvalité) som använder objektiv värderingsteknik.

**PESQ** (Perceptual Evaluation of Speech Quality) är en standard innehållande metoder för utvärdering av talkvalité. Algoritmen ger ett värde mellan -0,5 och 4,5 och ger en indikation på förväntat MOS-värde.

**STOI** (Short-Time Objective Intelligibility Measure) är en algoritm som beräknar taltydligheten genom att jämföra referenssignal med en bearbetad signal. Möjliga värde är mellan 0 och 1 där 1 är betyder "utmärkt".

**Vitt brus** (white noise) är brus vars frekvens- och energispektrum är konstant och oberoende av frekvensen. Signalenergin för en konstant bandbredd (centrerad kring en frekvens  $f_0$ ) ändras inte om  $f_0$  varierar. Namnet kommer från associationer till vitt ljus, där det finns lika mängd av alla färger.

## 8 Källförteckning

Audacity. (2019). Free, open source, cross-platform audio software. Hämtad 2019-03-04 från <https://www.audacityteam.org/>

Bareham, J. R. (2018). *Electroacoustic Measurements on Personal Communication Devices* [PowerPoint-presentation]. Hämtad 2018-11-09 från Brüel & Kjær University in Nærum, Danmark

Behringer. (2019). XENYX X1222USB. Hämtad 2019-02-18 från <https://www.musictribe.com/Categories/Behringer/Mixers/Analog/X1222USB/p/P0A0I>

Benesty, J., Sondhi, M. M., & Huang, Y. (Eds.). (2007). Springer handbook of speech processing. Springer.

Brandt, A. (2000). Ljud- och vibrationsanalys. 1. Täby: Saven EduTech.

*Brottsbalk* (SFS 1962:700). Hämtad från Riksdagens webbplats: [https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/brottsbalk-1962700\\_sfs-1962-700](https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/brottsbalk-1962700_sfs-1962-700)

Brüel & Kjær. (n.d). Hand-held analyzer Type 2250-S. Hämtad 2019-02-20 från <https://www.bksv.com/en/products/sound-and-vibration-meters/sound-level-meters-and-vibration-meters/Type-2250-S>

Calvert, J.B. (2003). Microphones. Hämtad 2018-11-10 från <http://mysite.du.edu/~jcalvert/tech/microph.htm>

Carter, B. & Mancini, R. (2009). *Op Amps for Everyone [Elektronisk resurs]*. (3rd ed.) Burlington: Elsevier.

Cigada, A., Ripamonti, F., & Vanali, M. (2007). The delay & sum algorithm applied to microphone array measurements: numerical analysis and experimental validation. *Mechanical Systems and Signal Processing*, 21(6), 2645-2664.

Decibelmätare. (2019). Ljudnivå. Hämtad 2019-01-10 från <https://decibelmatare.se/ljudniva/>

Fraden, J. (2016). Handbook of Modern Sensors Physics, Designs, and Applications. (5th ed.) Cham: Springer International Publishing.

Freeman, R. L. (2005). Fundamentals of telecommunications (Vol. 92). John Wiley & Sons.

Genelec. (n.d.a). 1037C Studio Monitor. Hämtad 2019-02-20 från <https://www.genelec.com/support-technology/previous-models/1037c-studio-monitor>

Genelec. (n.d.b). 8020C Studio Monitor. Hämtad 2019-02-20 från <https://www.genelec.com/studio-monitors/8000-series-studio-monitors/8020c-studio-monitor>

Genelec. (2013). Data sheet 8020C. Hämtad 2019-02-17 från [https://www.genelec.com/sites/default/files/media/Studio%20monitors/8000%20Series%20Studio%20Monitors/8020C/genelec\\_8020c\\_datasheet.pdf](https://www.genelec.com/sites/default/files/media/Studio%20monitors/8000%20Series%20Studio%20Monitors/8020C/genelec_8020c_datasheet.pdf)

GRAS Sound & Vibration. (n.d.a). GRAS 44AA Mouth Simulator according to ITU-T Rec. P51 with built-in power amplifier. Hämtad 2019-04-01 från <http://www.gras.dk/products/product/280-44AA.html>

GRAS Sound & Vibration. (n.d.b). GRAS 26CA. Hämtad 2019-04-01 från <https://www.gras.dk/products/preamplifiers-for-microphone-cartridge/constant-current-power-ccp/product/203-26ca>

Greenpeace. (2019). Elektronik. Hämtad 2019-02-24 från <https://www.greenpeace.org/archive-sweden/se/miljogifter/Elektronik/>

Grimsberg, M. (2005). Börja med Matlab (2.uppl.). Lund: KFS.

Götahälsan. (2016). Hörselnedsättning. Hämtad 2018-11-12 från <http://www.gotahalsan.se/sv/horselnedsattning/>

Hald, J & Christenson, J. J. (2002). A class of optimal broadband phased array geometries designed for easy construction. Naerum: Brüel & Kjaer Sound & Vibration Measurement A/S.

Håkansson, O. & Kjellander, P. (1991). The construction and placement of a microphone array for automotive use. Examensarbete.

IEEE Subcommittee on Subjective Measurements. (1969). *IEEE Transactions on Audio and Electroacoustics*, vol 17, 225-246.

International Telecommunication Union. (2007). P.862.3 : Application guide for objective quality measurement based on Recommendations P.862, P.862.1 and P.862.2. Hämtad 2019-01-17 från <https://www.itu.int/rec/T-REC-P.862.3-200711-I/en>

International Telecommunication Union. (2018). P.862 : Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. Hämtad 2019-01-17 från <https://www.itu.int/rec/t-rec-p.862>

Kjellson, A. (2014). Sound Source Localization and Beamforming for Teleconferencing Solutions.

Kondo, K. (2012). *Subjective Quality Measurement of Speech [Elektronisk resurs] Its Evaluation, Estimation and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Lai, C. C., Nordholm, S. E., & Leung, Y. H. (2017). A Study Into the Design of Steerable Microphone Arrays. Springer Singapore.

Loizou, P.C. (2013). *Speech enhancement: theory and practice*. (2. ed.) Boca Raton, Fla.: CRC Press.

MathWorks. (2011). PESQ MATLAB Wrapper. Hämtad 2019-01-17 från <https://www.mathworks.com/matlabcentral/fileexchange/33820-pesq-matlab-wrapper>

MathWorks. (2019). Beamforming Overview. Hämtad 2019-03-05 från [https://www.mathworks.com/help/phased/ug/beamforming-concepts.html?searchHighlight=beamforming&s\\_tid=doc\\_srchtile](https://www.mathworks.com/help/phased/ug/beamforming-concepts.html?searchHighlight=beamforming&s_tid=doc_srchtile)

Maxim Integrated. (2019). MAX9860EVKIT. Hämtad 2019-04-04 från <https://www.maximintegrated.com/en/products/analog/audio/MAX9860EVKIT.html>

McCowan, I. (2001). Microphone arrays: A tutorial. Queensland University, Australia, 1-38.

Morcelli, A., & Widder, J. (2014, 14 Maj). Basic principles of MEMS microphones. *EDN Network*. Hämtad från <https://www.edn.com/design/analog/4430264/Basic-principles-of-MEMS-microphones->

Nisula, J., & Krill, S. (2015). Acoustic Solutions for Door Station.

Postema, L., & Davidsson, J. (2019). Beamforming and Blind Signal Separation for Far-field Voice Capture using a Microphone Array.

RISE Research Institutes of Sweden. (2017). Kalibrering av ljudnivåmätare. Hämtad 2019-02-20 från [https://www.sp.se/sv/index/services/calibration\\_sound/calibration\\_sound/Sidor/default.aspx](https://www.sp.se/sv/index/services/calibration_sound/calibration_sound/Sidor/default.aspx)

Sseed Studio. (2019a). ReSpeaker Mic Array v2.0. Hämtad 2019-04-04 från <https://www.sseedstudio.com/ReSpeaker-Mic-Array-v2-0-p-3053.html>

Sseed Studio. (2019b). ReSpeaker Core v2.0. Hämtad 2019-04-04 från <https://www.sseedstudio.com/category/Development-Platforms-c-1002/single-board-computer-c-950/category/Speech-Recognition-c-44/ReSpeaker-Core-v2-0-p-3039.html>

Shen, L., Chu, Z., Yang, Y., & Wang, G. (2017). 2348. A new beamforming microphones array with acoustic insulation baffle. *Journal of Vibroengineering*, 19(1).

Shure. (2019). SRH940 Professional Reference Headphones. Hämtad 2019-02-20 från <https://www.shure.com/americas/products/headphones/srh940-professional-reference-headphones>

STMicroelectronics. (2019). MEMS Microphones. Hämtad 2019-04-01 från <https://www.st.com/en/mems-and-sensors/mems-microphones.html>

Strong, J. (2002). Digital ljudinspelning i ett nötskal: [goda råd om inspelningsutrustning, MIDI, mikrofoner, effekter m.m. : en referenshandbok för oss andra!]. Sundbyberg: Pagina.

Studiefrämjandet. (n.d.). Mikrofoner. Hämtad 2018-11-19 från <https://www.studieframjandet.se/musikakuten/meny-4/att-rigga-och-spela-in/mikroner/>

Taal, C. H., Hendriks, R. C., Heusdens, R., & Jensen, J. (2011). An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7), 2125-2136.

Texas Instruments. (2019). Audio Pre-processing System Reference Design for VoiceBased Applications Using C5517. Hämtad 2019-04-04 från <http://www.ti.com/lit/ug/tiducy1c/tiducy1c.pdf>

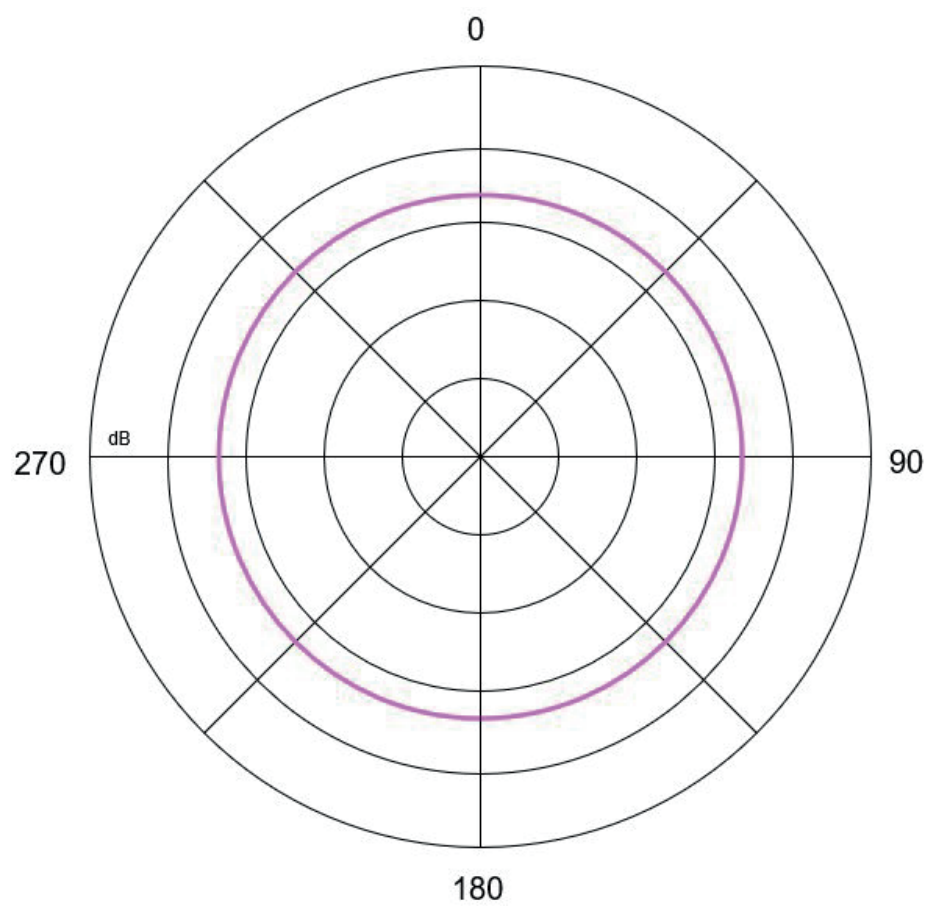
RME. (n.d.). Fireface 802. Hämtad 2019-02-18 från [http://www.rme-audio.de/en/products/fireface\\_802.php](http://www.rme-audio.de/en/products/fireface_802.php)

Yamaha. (n.d.). MGP Series. Hämtad 2019-02-18 från <https://asia-latinamerica-mea.yamaha.com/en/products/proaudio/mixers/mgp/index.html#product-tabs>

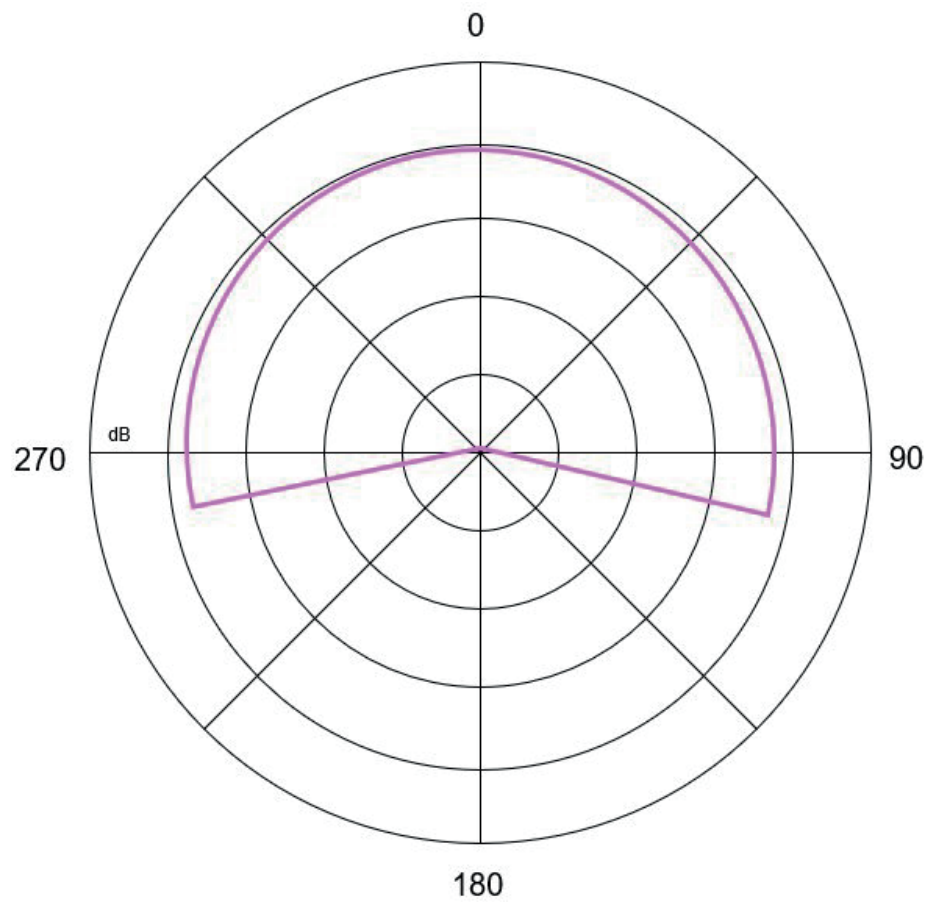
Ögren, M. (1996). *The design of an array microphone*. Examensarbete. Borås.

## Appendix A: Polär diagram

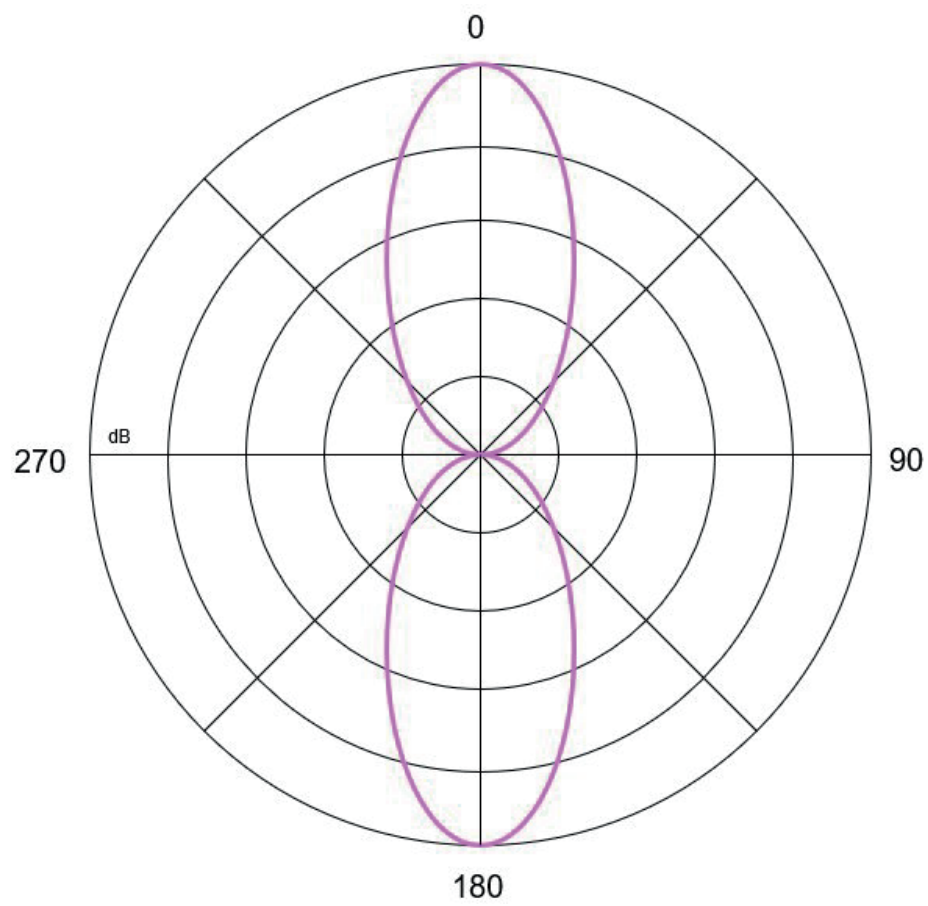
A.1: Skiss över ett polär diagram som representerar omnidirektionella mikrofoners ljudupptagning.



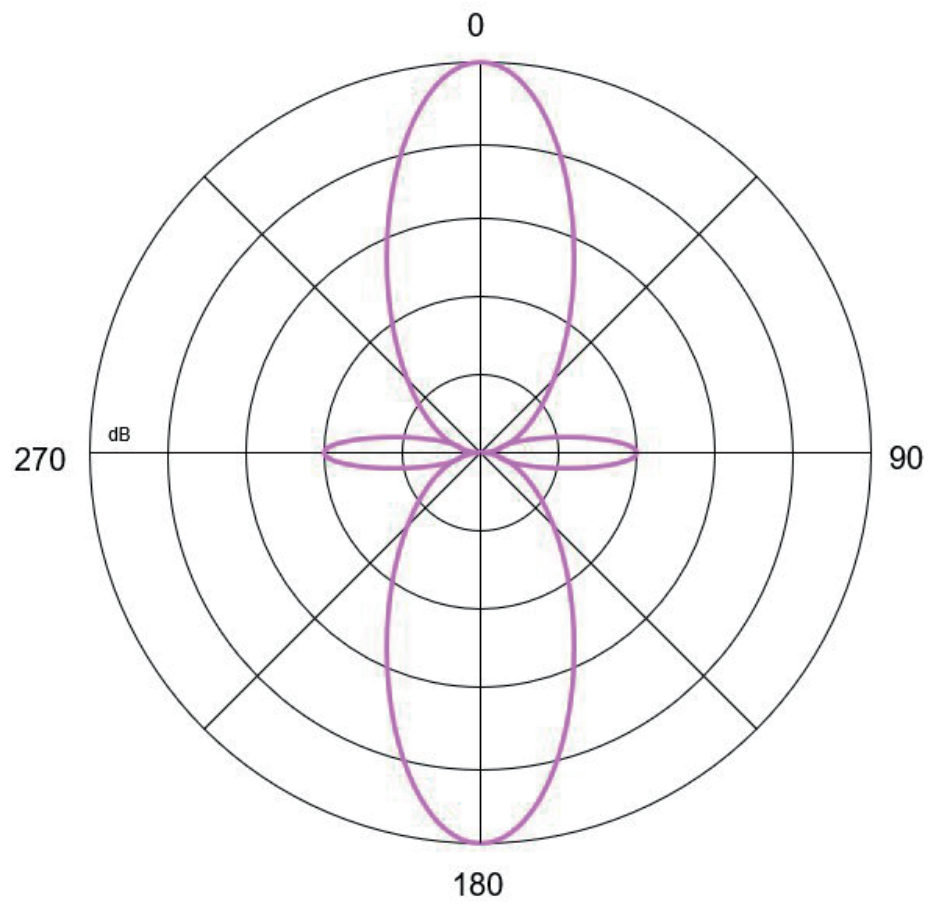
A.2: Skiss över ett polärdiagram som representerar unidirektionella mikrofoners ljudupptagning.



A.3: Skiss över ett polärdiagram som representerar åttakopplade mikrofoners ljudupptagning.



A.4: Skiss över ett polärdiagram som representerar åttakopplade mikrofoners ljudupptagning med sidolober.



## Appendix B: MATLAB-kod för bearbetning av ljudfiler

B.1: MATLAB-kod för beamforming av två ljudfiler.

### **tva\_filer.m**

```
[t1,Fs1]=audioread('Fil_mic_1.wav');
[t2,Fs2]=audioread('Fil_mic_2.wav');
if Fs1 ~= Fs2
    error("Not the same sample rate!")
end
fs = Fs1;
all_lengths=[length(t1), length(t2)];
shortest_samples=min(all_lengths);
t_duration = shortest_samples / fs;
%% Prepare for beamforming
ang_dft = [-180; 37];
sigArray = [t1,t2];
c=340;
% antal mikrofoner
Nele = 2;
angSteer = ang_dft;
microphone = phased.OmnidirectionalMicrophoneElement('FrequencyRange',[20 20e3]);
ula = phased.ULA(Nele,0.045,'Element',microphone);
%% beamforming - Time Delay
beamformer = phased.TimeDelayBeamformer('SensorArray',ula,...
    'SampleRate',fs,'Direction',angSteer,'PropagationSpeed',c);
yTimeDelay = beamformer(sigArray);
t = 0:1/fs:t_duration-1/fs;
plot(t,sigArray(:,2),'r',t,yTimeDelay)
xlabel('Time (sec)')
ylabel('Amplitude')
legend('Original (T3)','Beamformed')
%% Play the beamformed audio
audioWriter = audioDeviceWriter('SampleRate',fs, ...
    'SupportVariableSizeInput', true);
play(audioWriter,yTimeDelay);
```

```
filename = 'Fil_namn.wav';  
audiowrite(filename,yTimeDelay,fs);
```

B.2: MATLAB-kod för beamforming av tre ljudfiler.

**tre\_filer.m**

```
[t1,Fs1]=audioread('Fil_mic_1.wav');
[t2,Fs2]=audioread('Fil_mic_2.wav');
[t3,Fs3]=audioread('Fil_mic_3.wav');
if Fs1 ~= Fs2 || Fs2 ~= Fs3
    error("Not the same sample rate!")
end
fs = Fs1;
all_lengths=[length(t1), length(t2), length(t3)];
shortest_samples=min(all_lengths);
t_duration = shortest_samples / fs;
%% Prepare for beamforming
ang_dft = [-180; 37];
sigArray = [t1,t2,t3];
c=340;
% antal mikrofoner
Nele = 3;
angSteer = ang_dft;
microphone = phased.OmnidirectionalMicrophoneElement('FrequencyRange',[20 20e3]);
ula = phased.ULA(Nele,0.045,'Element',microphone);
%% beamforming - Time Delay
beamformer = phased.TimeDelayBeamformer('SensorArray',ula,...
    'SampleRate',fs,'Direction',angSteer,'PropagationSpeed',c);
yTimeDelay = beamformer(sigArray);
t = 0:1/fs:t_duration-1/fs;
plot(t,sigArray(:,3),'r',t,yTimeDelay)
xlabel('Time (sec)')
ylabel('Amplitude')
legend('Original (T3)','Beamformed')
audioWriter = audioDeviceWriter('SampleRate',fs, ...
    'SupportVariableSizeInput', true);
play(audioWriter,yTimeDelay);
filename = 'Fil_namn.wav';
audiowrite(filename,yTimeDelay,fs);
```

B.3: MATLAB-kod för beamforming av fyra ljudfiler.

**fyra\_filer.m**

```
[t1,Fs1]=audioread(Fil_mic_1.wav');
[t2,Fs2]=audioread(Fil_mic_2.wav');
[t3,Fs3]=audioread(Fil_mic_3.wav');
[t4,Fs4]=audioread(Fil_mic_4.wav');
if Fs1 ~= Fs2 || Fs2 ~= Fs3 || Fs3 ~= Fs4
    error("Not the same sample rate!")
end
fs = Fs1;
all_lengths=[length(t1), length(t2), length(t3), length(t4)];
shortest_samples=min(all_lengths);
t_duration = shortest_samples / fs;
ang_dft = [-180; 37];
sigArray = [t1,t2,t3,t4];
c=340;
% antal mikrofoner
Nele = 4;
angSteer = ang_dft;
microphone = phased.OmnidirectionalMicrophoneElement('FrequencyRange',[20 20e3]);
ula = phased.ULA(Nele,0.045,'Element',microphone);
%% beamforming - Time Delay
beamformer = phased.TimeDelayBeamformer('SensorArray',ula,...
    'SampleRate',fs,'Direction',angSteer,'PropagationSpeed',c);
yTimeDelay = beamformer(sigArray);
t = 0:1/fs:t_duration-1/fs;
plot(t,sigArray(:,3),'r',t,yTimeDelay)
xlabel('Time (sec)')
ylabel('Amplitude')
legend('Original (T3)','Beamformed')
audioWriter = audioDeviceWriter('SampleRate',fs, ...
    'SupportVariableSizeInput', true);
%play(audioWriter,yTimeDelay);
filename = 'Fil_namn.wav';
audiowrite(filename,yTimeDelay,fs);
```

B.4: MATLAB-kod för korrelering av ljudfiler.

#### **Correlation.m**

```
load relatedsig
[t1, Fs] = audioread('Renttal.wav');
[t2, Fs] = audioread('Tal+Noise.wav');
ax(1) = subplot(3,1,1);
plot(t1)
ylabel('s_1')
axis tight
ax(2) = subplot(3,1,2);
plot(t2)
ylabel('s_2')
axis tight
linkaxes(ax,'x')

[C21,lag21] = xcorr(t2,t1);
C21 = C21/max(C21);
[M21,I21] = max(C21);
t21 = lag21(I21);
subplot(3,1,1)
plot(lag21,C21,[t21 t21],[-0.5 1],':r')
text(t21+100,0.5,['Lag: ' int2str(t21)])
ylabel('C_{21}')
axis tight
title('Cross-Correlations')
subplot(3,1,2)
ylabel('C_{31}')
axis tight
```

B.5: MATLAB-kod för beräkning av STOI-värde.

**main.Stoi.m**

```
[x, Fs] = audioread('Ljudfil ren.wav');  
[z, Fs] = audioread('Ljudfil med brus.wav');  
  
res = stoi(x, z, Fs)
```

**stoi.m**

```
function d = stoi(x, y, fs_signal)  
if length(x)~=length(y)  
    error('x and y should have the same length');  
end  
% initialization  
x      = x(:);  
y      = y(:);  
fs     = 10000;  
N_frame = 256;  
K      = 512;  
J      = 15;  
mn     = 150;  
H      = thirdoct(fs, K, J, mn);  
N      = 30;  
Beta   = -15;  
dyn_range = 40;  
  
if fs_signal ~= fs  
    x = resample(x, fs, fs_signal);  
    y = resample(y, fs, fs_signal);  
end  
  
[x y] = removeSilentFrames(x, y, dyn_range, N_frame, N_frame/2);  
  
x_hat = stdft(x, N_frame, N_frame/2, K);  
y_hat = stdft(y, N_frame, N_frame/2, K);  
x_hat = x_hat(:, 1:(K/2+1)).';
```

```

y_hat = y_hat(:, 1:(K/2+1)).';
X      = zeros(J, size(x_hat, 2));
Y      = zeros(J, size(y_hat, 2));
for i = 1:size(x_hat, 2)
    X(:, i) = sqrt(H*abs(x_hat(:, i)).^2);
    Y(:, i) = sqrt(H*abs(y_hat(:, i)).^2);
End

d_interm = zeros(J, length(N:size(X, 2)));
c         = 10^(-Beta/20);
for m = N:size(X, 2)
    X_seg = X(:, (m-N+1):m);
    Y_seg = Y(:, (m-N+1):m);
    alpha = sqrt(sum(X_seg.^2, 2)/sum(Y_seg.^2, 2));
    aY_seg = Y_seg.*repmat(alpha, [1 N]);
    Y_prime = min(aY_seg(j, :), X_seg(j, :)+X_seg(j, :)*c);
    d_interm(j, m-N+1) = taa_corr(X_seg(j, :).', Y_prime(:));
end

end

d = mean(d_interm(:));
%%%
function [A cf] = thirdoct(fs, N_fft, numBands, mn)
f      = linspace(0, fs, N_fft+1);
f      = f(1:(N_fft/2+1));
k      = 0:(numBands-1);
cf     = 2.^(k/3)*mn;
fl     = sqrt((2.^(k/3)*mn).*2.^((k-1)/3)*mn);
fr     = sqrt((2.^(k/3)*mn).*2.^((k+1)/3)*mn);
A      = zeros(numBands, length(f));
for i = 1:(length(cf))
    [a b]      = min((f-fl(i)).^2);
    fl(i)      = f(b);
    fl_ii      = b;
[a b]         = min((f-fr(i)).^2);
    fr(i)      = f(b);
    fr_ii      = b;

```

```

        A(i,fl_ii:(fr_ii-1)) = 1;
    end
    rnk      = sum(A, 2);
    numBands = find((rnk(2:end)>=rnk(1:(end-1))) & (rnk(2:end)~=0)~=0, 1, 'last')+1;
    A        = A(1:numBands, :);
    cf       = cf(1:numBands);
    %%
    function x_stdft = stdft(x, N, K, N_fft)
    frames = 1:K:(length(x)-N);
    x_stdft = zeros(length(frames), N_fft);
    w        = hanning(N);
    x        = x(:);
    for i = 1:length(frames)
        ii      = frames(i):(frames(i)+N-1);
        x_stdft(i, :) = fft(x(ii).*w, N_fft);
    end
    %%
    function [x_sil y_sil] = removeSilentFrames(x, y, range, N, K)
    x        = x(:);
    y        = y(:);
    frames = 1:K:(length(x)-N);
    w        = hanning(N);
    msk      = zeros(size(frames));
    for j = 1:length(frames)
        jj      = frames(j):(frames(j)+N-1);
        msk(j) = 20*log10(norm(x(jj).*w)./sqrt(N));
    end
    msk      = (msk-max(msk)+range)>0;
    count = 1;
    x_sil = zeros(size(x));
    y_sil = zeros(size(y));
    for j = 1:length(frames)
        if msk(j)
            jj_i      = frames(j):(frames(j)+N-1);
            jj_o      = frames(count):(frames(count)+N-1);
            x_sil(jj_o) = x_sil(jj_o) + x(jj_i).*w;
            y_sil(jj_o) = y_sil(jj_o) + y(jj_i).*w;
        end
        count = count + 1;
    end
end

```

```

        count = count+1;
    end
end
x_sil = x_sil(1:jj_o(end));
y_sil = y_sil(1:jj_o(end));
%%
function rho = taa_corr(x, y)
    xn = x-mean(x);
    xn = xn/sqrt(sum(xn.^2));
    yn = y-mean(y);
    yn = yn/sqrt(sum(yn.^2));
    rho = sum(xn.*yn);

```

B.6: MATLAB-kod för beräkning av PESQ-värde.

<https://se.mathworks.com/matlabcentral/fileexchange/33820-pesq-matlab-wrapper>

points at

<https://www.itu.int/rec/T-REC-P.862-200511-I!Amd2/en>

The changes made to the code base:

```
--- /Software/orig/P862_annex_A_2005_CD/source/pesqpar.h
```

```
+++ /Software/fixed/P862_annex_A_2005_CD/source/pesqpar.h
```

```
@@@ -122,6 +122,7 @@@
```

```
long InIIR_Nsos;
```

```
+#define LINIIR 60
```

```
long InIIR_Nsos_8k = 8L;
```

```
float InIIR_Hsos_8k[LINIIR] =
```

```
{ 0.885535424f,    -0.885535424f, 0.000000000f, -0.771070709f, 0.000000000f,
```

B.7: MATLAB-kod för generering av grafiskt spektrum.

### **Sound\_Analysis.m**

```
clear, clc, close all

[x, fs] = audioread('Fil_namn.wav');
x = x(:, 1);
N = length(x);
t = (0:N-1)/fs;
figure(2)
spectrogram(x, 1024, 3/4*1024, [], fs, 'yaxis')
box on
set(gca, 'FontName', 'Times New Roman', 'FontSize', 14)
xlabel('Time, s')
ylabel('Frequency, Hz')
title('Spectrogram of the signal')
h = colorbar;
set(h, 'FontName', 'Times New Roman', 'FontSize', 14)
ylabel(h, 'Magnitude, dB')
% spectral analysis
w = hanning(N, 'periodic');
[X, f] = periodogram(x, w, N, fs, 'power');
X = 20*log10(sqrt(X)*sqrt(2));
[Rx, lags] = xcorr(x, 'coeff');
d = lags/fs;
maxval = max(x);
minval = min(x);
disp(['Max value = ' num2str(maxval)])
disp(['Min value = ' num2str(minval)])
u = mean(x);
s = std(x);
disp(['Mean value = ' num2str(u)])
disp(['RMS value = ' num2str(s)])
D = 20*log10(maxval/min(abs(nonzeros(x))));
disp(['Dynamic range D = ' num2str(D) ' dB'])
Q = 20*log10(maxval/s);
disp(['Crest factor Q = ' num2str(Q) ' dB'])
```

```
ind = find(Rx>0.05, 1, 'last');  
RT = (ind-N)/fs;  
disp(['Autocorrelation time = ' num2str(RT) ' s'])  
commandwindow
```

## Appendix C: Blankett för poängsättning vid MOS-test

MOS Rating Scores

| Rating | Speech Quality | Level of Distortion                |
|--------|----------------|------------------------------------|
| 5      | Excellent      | Imperceptible                      |
| 4      | Good           | Just perceptible, but not annoying |
| 3      | Fair           | Perceptible and slightly annoying  |
| 2      | Poor           | Annoying, but not objectionable    |
| 1      | Bad            | Very annoying and objectionable    |

|                     |  |
|---------------------|--|
| Poäng för ljudfil 1 |  |
| Poäng för ljudfil 2 |  |
| Poäng för ljudfil 3 |  |
| Poäng för ljudfil 4 |  |
| Poäng för ljudfil 5 |  |
| Poäng för ljudfil 6 |  |
| Poäng för ljudfil 7 |  |
| Poäng för ljudfil 8 |  |
| Poäng för ljudfil 9 |  |

|                      |  |
|----------------------|--|
| Poäng för ljudfil 10 |  |
| Poäng för ljudfil 11 |  |
| Poäng för ljudfil 12 |  |
| Poäng för ljudfil 13 |  |

TACK för hjälpen!

## Appendix D: Innebörden av ljudfilernas numrering vid MOS-test

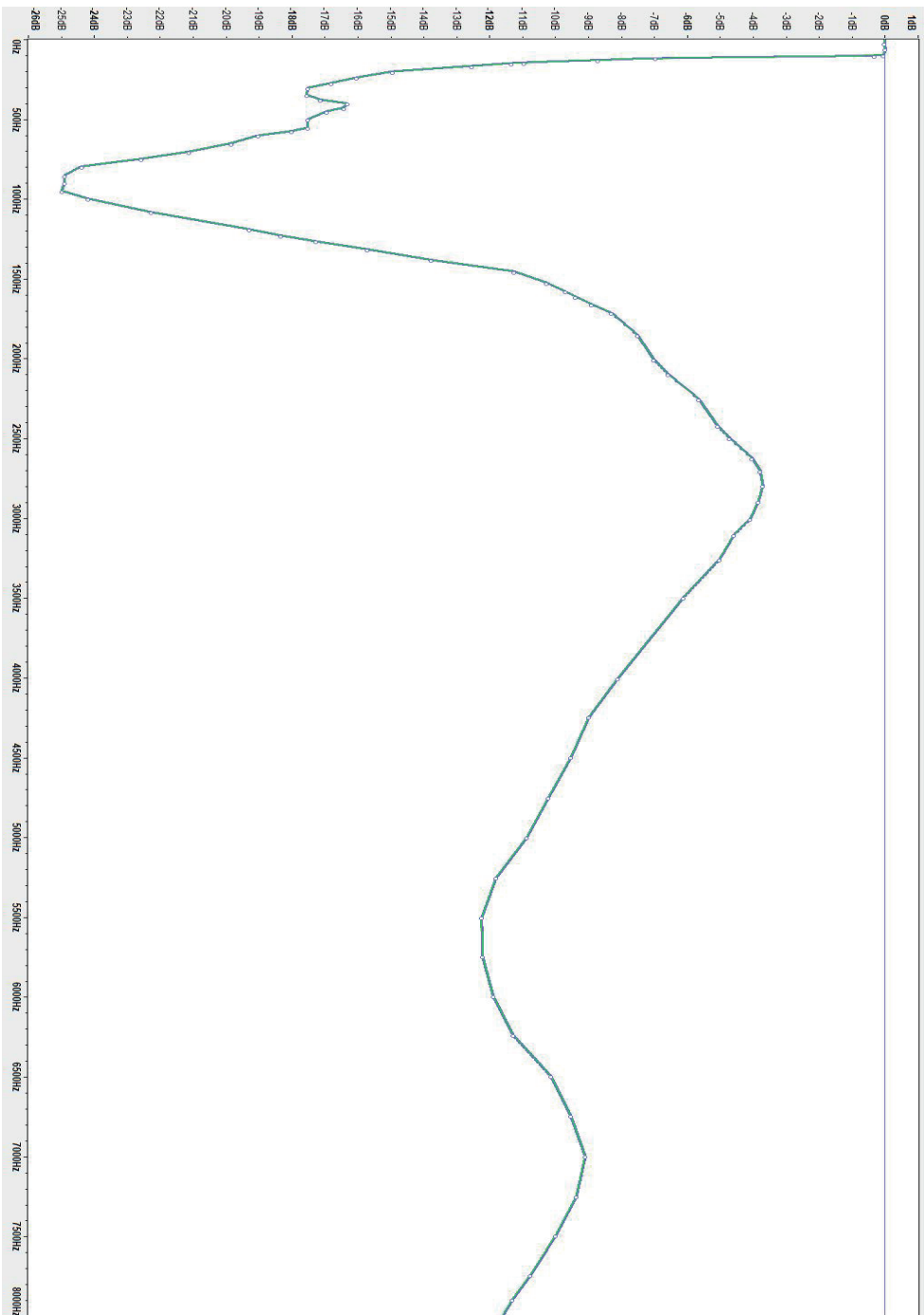
|            |                               |
|------------|-------------------------------|
| Ljudfil 1  | Svanhals                      |
| Ljudfil 2  | Typ A, front: 2 mikrofoner    |
| Ljudfil 3  | Typ A, front: 3 mikrofoner    |
| Ljudfil 4  | Typ A, front: 4 mikrofoner    |
| Ljudfil 5  | Typ A, ovansida: 2 mikrofoner |
| Ljudfil 6  | Typ A, ovansida: 3 mikrofoner |
| Ljudfil 7  | Typ A, ovansida: 4 mikrofoner |
| Ljudfil 8  | Typ B, front: 2 mikrofoner    |
| Ljudfil 9  | Typ B, front: 3 mikrofoner    |
| Ljudfil 10 | Typ B, front: 4 mikrofoner    |
| Ljudfil 11 | Typ B, ovansida: 2 mikrofoner |
| Ljudfil 12 | Typ B, ovansida: 3 mikrofoner |
| Ljudfil 13 | Typ B, ovansida: 4 mikrofoner |

## Appendix E: Individuella svar i MOS-testet

|             | 1                   | 2       | 3       | 4       | 5          | 6          | 7          | 8                   | 9       | 10      | 11         | 12         | 13         |
|-------------|---------------------|---------|---------|---------|------------|------------|------------|---------------------|---------|---------|------------|------------|------------|
|             | Mikrofoner av typ A |         |         |         |            |            |            | Mikrofoner av typ B |         |         |            |            |            |
| Lyssnare nr | svanhals            | front 2 | front 3 | front 4 | ovansida 2 | ovansida 3 | ovansida 4 | front 2             | front 3 | front 4 | ovansida 2 | ovansida 3 | ovansida 4 |
|             | 5                   | 1       | 1       | 2       | 1          | 2          | 2          | 3                   | 3       | 4       | 4          | 5          | 5          |
|             | 3                   | 2       | 2       | 3       | 3          | 2          | 2          | 3                   | 3       | 3       | 3          | 4          | 4          |
|             | 4                   | 2       | 2       | 3       | 2          | 3          | 3          | 4                   | 4       | 3       | 3          | 4          | 3          |
|             | 3                   | 1       | 1       | 2       | 1          | 1          | 2          | 1                   | 2       | 2       | 3          | 2          | 1          |
|             | 2                   | 1       | 2       | 2       | 1          | 1          | 2          | 3                   | 3       | 3       | 3          | 3          | 3          |
|             | 3                   | 1       | 2       | 2       | 1          | 1          | 2          | 3                   | 3       | 4       | 3          | 3          | 3          |
|             | 3                   | 2       | 2       | 3       | 2          | 2          | 2          | 4                   | 4       | 4       | 4          | 3          | 3          |
|             | 5                   | 2       | 2       | 3       | 2          | 2          | 3          | 3                   | 3       | 4       | 3          | 4          | 4          |
|             | 3                   | 2       | 2       | 2       | 1          | 2          | 2          | 4                   | 4       | 4       | 4          | 4          | 3          |
|             | 2                   | 3       | 3       | 3       | 3          | 3          | 3          | 2                   | 2       | 2       | 3          | 3          | 3          |
|             | 3                   | 2       | 2       | 3       | 2          | 3          | 3          | 4                   | 3       | 4       | 3          | 4          | 4          |
|             | 2                   | 1       | 2       | 2       | 1          | 2          | 2          | 3                   | 3       | 3       | 4          | 3          | 3          |
|             | 3                   | 2       | 3       | 3       | 2          | 3          | 3          | 4                   | 4       | 4       | 4          | 3          | 3          |

|                    |     |      |      |     |      |    |     |      |     |      |      |     |     |
|--------------------|-----|------|------|-----|------|----|-----|------|-----|------|------|-----|-----|
|                    | 2   | 1    | 1    | 1   | 1    | 2  | 2   | 3    | 3   | 3    | 3    | 3   | 3   |
|                    | 2   | 1    | 1    | 1   | 1    | 1  | 1   | 3    | 3   | 3    | 3    | 3   | 3   |
|                    | 1   | 1    | 2    | 2   | 1    | 2  | 3   | 4    | 5   | 3    | 3    | 2   | 3   |
|                    | 2   | 1    | 1    | 2   | 1    | 1  | 2   | 4    | 4   | 4    | 3    | 3   | 3   |
|                    | 3   | 1    | 1    | 2   | 1    | 2  | 2   | 4    | 4   | 4    | 3    | 4   | 4   |
|                    | 3   | 3    | 4    | 4   | 4    | 3  | 4   | 5    | 5   | 5    | 5    | 5   | 5   |
|                    | 2   | 1    | 3    | 3   | 2    | 2  | 3   | 3    | 3   | 3    | 3    | 3   | 3   |
| sam<br>manl<br>agt | 56  | 31   | 39   | 48  | 33   | 40 | 48  | 67   | 68  | 69   | 67   | 68  | 66  |
| Mede<br>lvärd<br>e | 2,8 | 1,55 | 1,95 | 2,4 | 1,65 | 2  | 2,4 | 3,35 | 3,4 | 3,45 | 3,35 | 3,4 | 3,3 |

## Appendix F: Grafisk visualisering av filter för förarbetning av ljudfil till munsimulator





**LUND**  
UNIVERSITY

Series of Bachelor's theses  
Department of Electrical and Information Technology  
LU/LTH-EIT 2019-693

<http://www.eit.lth.se>