Popular Summary

Imagine an action you regularly perform. As an example, consider turning around to face a friend who has called for your attention. In order to achieve this, you use the sound you heard to know which direction to turn to. If you instead wanted to pick up an object, you need to use a combination of your sense of sight and touch to first place your hand in the approximately right position and then grab with the right amount of force. In general, processing sensor data into useful information is key, regardless of the task at hand. While many types of sensors exist, each with their respective advantages and disadvantages, the primary focus of this thesis is how to use sound to gain information about our surroundings. More precisely, we focus on finding the location of various sound sources, microphones or walls, see Figure P.1.

How do we use sound recordings to infer where objects are? An illustrative example is a thunder strike. You first see a flash of light, three seconds later you hear the sound of thunder. Since the sound traveled for three seconds and the speed of sound is around 340 m/s, you can conclude that the distance to the strike is about one kilometer. Similarly, if the same sound is picked up by two microphones, the sound will arrive at the closer microphone first. In this way, we can process the recorded sound to obtain information on the relative positions of microphones and sound sources.

After the signal processing is completed, we know simple geometric relationships such as the distances between pairs of objects. These basic measurements can then be refined into more high-level knowledge, such as the 3D position of each object. This could for instance be finding the position of a sound source using the sound recorded from a few microphones of known position. A more complex and less well-known problem is self-calibration. Essentially, even without prior knowledge of the position of sound sources and microphones, we can simultaneously estimate all the locations using only the recorded sound.

These types of problems are challenging because they involve real-world data. If we consider a robot arm automating a task in a factory, it can perform difficult tasks with a high degree of efficiency. However, it can only operate in a very controlled environment. If the conditions change, such as just moving the position of the robot arm slightly, it will completely stop the robot arm from performing its task. For the problems described above, it is common to not only have noisy measurements but also for some to be incorrect altogether. While the problems are still solvable, they require methods with robustness; how to achieve this is the subject of this thesis.

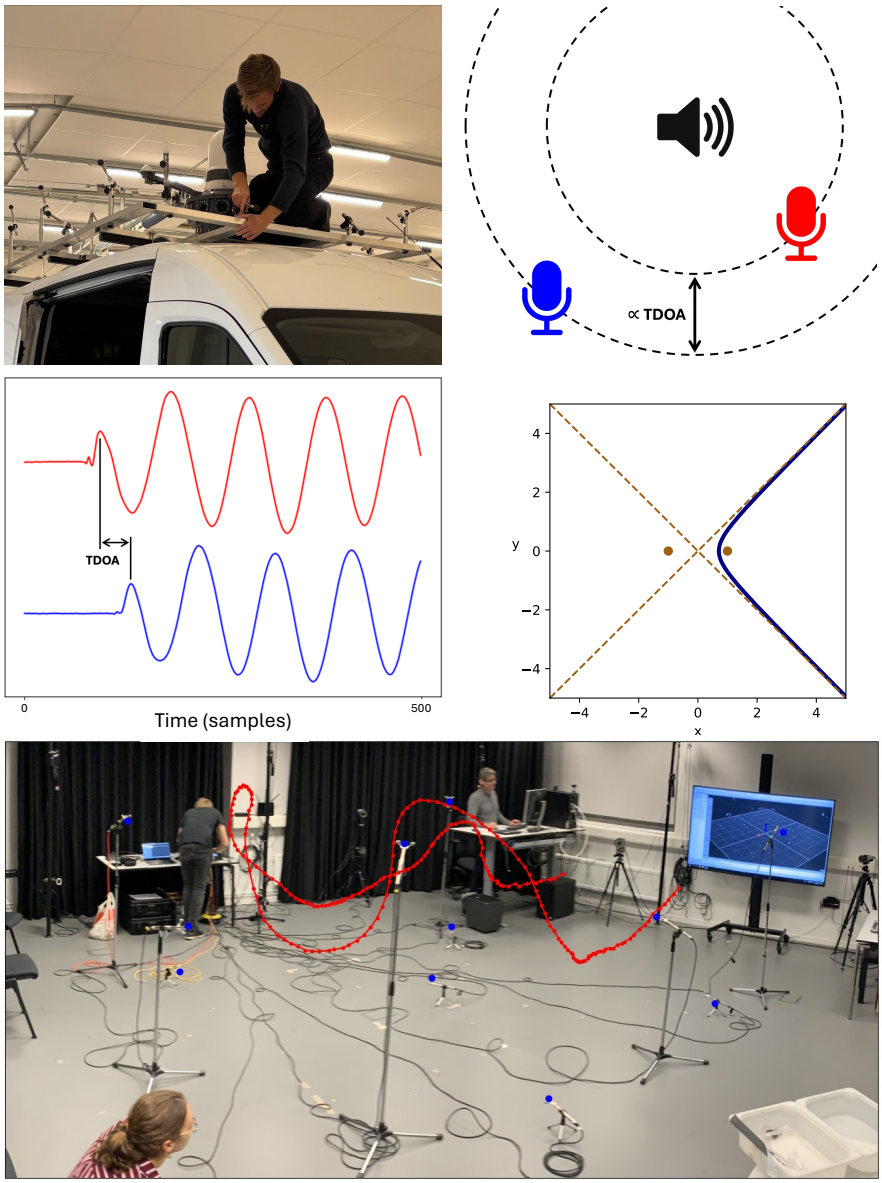

Figure P.1: (**top left**) A set of microphones with unknown positions. (**top right**) Using pairs of microphones, it is possible to measure the marked distance. (**center left**) Measuring is done by finding the corresponding time-delay in the recorded sound of the two microphones. (**center right**) This measurement reduces the possible positions of the sound source relative to the microphones. Possible locations are shown as the blue curve and the two microphones are the two bronze points. (**bottom**) Combining many of these measurements allows for the simultaneous estimation of the position of both microphones and speakers. Here the estimated microphones are marked as blue points and the red trajectory is the estimated path a speaker was moved along.

Populärvetenskaplig Sammanfattning

Föreställ dig en handling som du ofta utför. Som exempel kan vi tänka oss att du vänder dig mot en vän som har ropat på dig. För att veta åt vilket håll du ska vända dig använder du din hörsel. Om du istället vill plocka upp ett föremål behöver du använda en kombination av syn och känsel för att först placera handen ungefär rätt och sedan greppa med lagom kraft. Den gemensamma nämnaren är att du behöver använda dina sinnen för att interagera med din omvärld, det vill säga omvandla sensordata till användbar information. På samma sätt är det värdefullt att utveckla algoritmer som kan tolka sensordata, eftersom det skulle förbättra maskiners förmåga att interagera med omvärlden. Fokus för denna avhandling är hur ljud kan användas för att ta reda på information om vår omgivning. Mer specifikt fokuserar vi på att hitta positionerna för olika ljudkällor, mikrofoner eller väggar, se Figure P.2.

Hur går vi från ljud till att veta var saker befinner sig? Ett illustrativt exempel är ett blixtnedslag. Du ser först ett ljussken, och tre sekunder senare hör du åskan. Eftersom ljudet har färdats i tre sekunder och ljudets hastighet är ungefär 340 m/s kan du dra slutsatsen att avståndet till nedslaget är ungefär en kilometer. Om vi istället har två mikrofoner kan vi på liknande sätt lista ut att mikrofonen som är närmast ljudkällan kommer höra ljudet först. På detta sätt kan vi utföra signalbehandling för att gå från ljud till information om var mikrofoner och ljudkällor befinner sig.

När signalbehandlingen är klar har vi mätningar, till exempel avståndet mellan par av objekt. Dessa enkla mätningar kan sedan användas för att dra mer användbara slutsatser, till exempel den tredimensionella positionen för varje objekt. Beroende på vilken information man har till att börja med blir det olika geometriproblem att lösa. En typ av problem är att vi vet var våra mikrofoner är men vill identifiera positionen av en ljudkälla. Ett mer komplext och mindre välkänt problem är självkalibrering. Det innebär att utan förhandskunskap om var ljudkällor och mikrofoner befinner sig är det möjligt att skatta alla positioner samtidigt, endast baserat på det inspelade ljudet.

Dessa geometriska problem är utmanande eftersom de involverar riktig data. Om vi till exempel tänker oss en robotarm som automatiserar en uppgift i en fabrik, kan den utföra komplexa uppgifter med hög effektivitet. Den fungerar dock bara i en kontrollerad miljö. Om vi till exempel skulle flytta robotarmen en liten bit skulle den inte längre utföra sin uppgift, den saknar robusthet med andra ord. För de problem som beskrivits ovan är det vanligt att inte bara ha brusiga mätningar, utan även att vissa mätningar är direkt felaktiga. Även om problemen fortfarande går att lösa kräver de metoder med robusthet, hur detta kan uppnås är ämnet för denna avhandling.

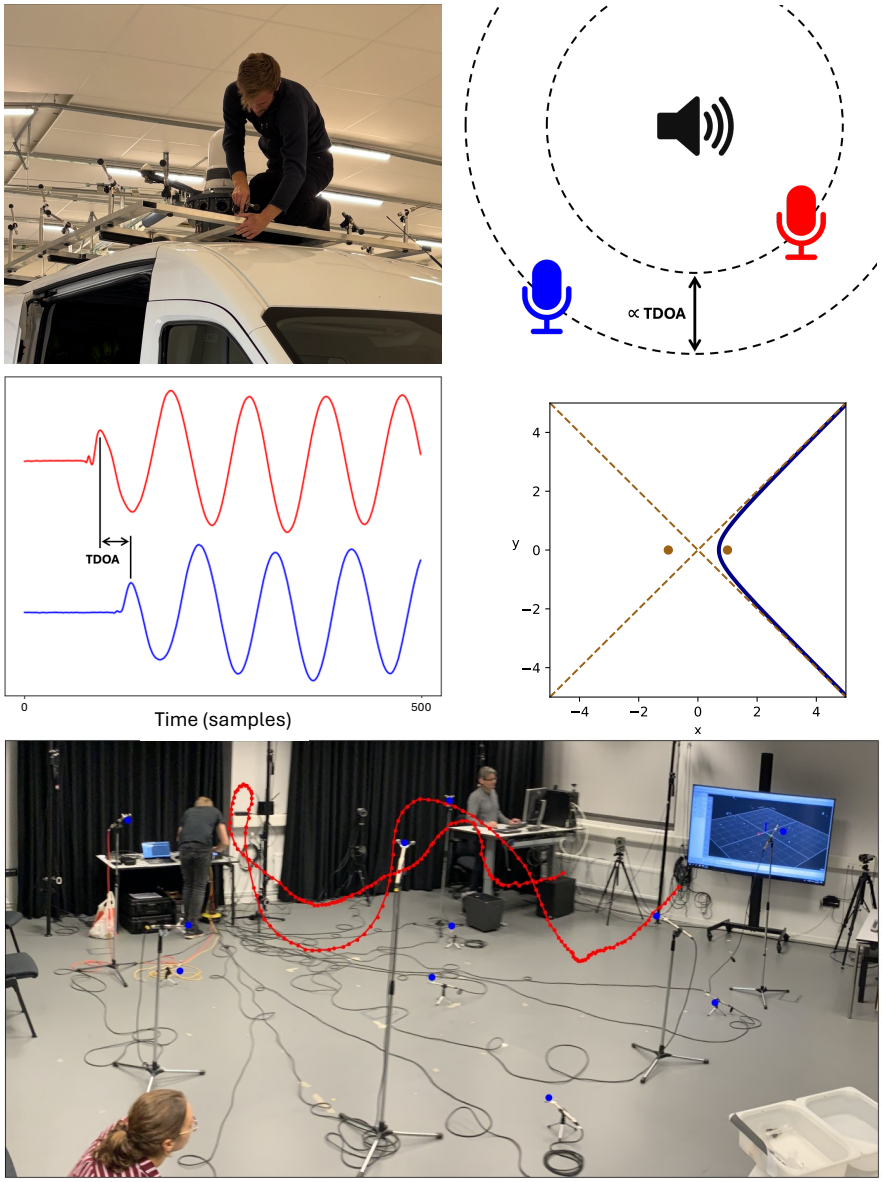

Figure P.2: (**övre vänster**) En samling av mikrofoner med okända positioner. (**övre höger**) Genom att använda par av mikrofoner är det möjligt att mäta det markerade avståndet. (**mitten vänster**) Detta görs genom att uppskatta tidsfördröjningen mellan att ljudet hörs i respektive av de två mikrofonerna. (**mitten höger**) På grund av denna mätning kan vi begränsa var ljudkällan kan befinna sig. Möjliga positioner för ljudkällan är markerat med blått och de två mikrofonerna är de bronsfärgade punkterna. (**nedre**) Genom att kombinera flera av dessa mätningar är det möjligt att samtidigt uppskatta positionerna för mikrofonerna och ljudkällorna. I bilden är de uppskattade mikrofonpositionerna markerade som blåa punkter och banan vi tror ljudkällan har rört sig längs är markerad med rött.